Origin and Evolution of Carboxysome Positioning Systems in Cyanobacteria

Joshua S. MacCready,[1] Joseph L. Basalla,[1] and Anthony G. Vecchiarelli *,[1]
[1]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI
*Corresponding author: E-mail: ave@umich.edu.
Associate editor: Julian Echave

Abstract

Carboxysomes are protein-based organelles that are essential for allowing cyanobacteria to fix CO_2. Previously, we identified a two-component system, McdAB, responsible for equidistantly positioning carboxysomes in the model cyanobacterium *Synechococcus elongatus* PCC 7942 (MacCready JS, Hakim P, Young EJ, Hu L, Liu J, Osteryoung KW, Vecchiarelli AG, Ducat DC. 2018. Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* 7:pii:e39723). McdA, a ParA-type ATPase, nonspecifically binds the nucleoid in the presence of ATP. McdB, a novel factor that directly binds carboxysomes, displaces McdA from the nucleoid. Removal of McdA from the nucleoid in the vicinity of carboxysomes by McdB causes a global break in McdA symmetry, and carboxysome motion occurs via a Brownian-ratchet-based mechanism toward the highest concentration of McdA. Despite the importance for cyanobacteria to properly position their carboxysomes, whether the McdAB system is widespread among cyanobacteria remains an open question. Here, we show that the McdAB system is widespread among β-cyanobacteria, often clustering with carboxysome-related components, and is absent in α-cyanobacteria. Moreover, we show that two distinct McdAB systems exist in β-cyanobacteria, with Type 2 systems being the most ancestral and abundant, and Type 1 systems, like that of *S. elongatus*, possibly being acquired more recently. Lastly, all McdB proteins share the sequence signatures of a protein capable of undergoing liquid–liquid phase separation. Indeed, we find that representatives of both McdB types undergo liquid–liquid phase separation in vitro, the first example of a ParA-type ATPase partner protein to exhibit this behavior. Our results have broader implications for understanding carboxysome evolution, biogenesis, homeostasis, and positioning in cyanobacteria.

Key words: **cyanobacteria, carboxysomes, subcellular organization, McdAB, ParAB.**

Introduction

The ability for cells to organize their interior is ubiquitous across all domains of life. In bacteria, the ParA/MinD family of ATPases has been primarily studied for their ability to segregate genetic cargos, such as chromosomes and plasmids (reviews in Baxter and Funnell 2014; Badrinarayanan et al. 2015). Less studied are ParA/MinD family members implicated in the positioning of diverse protein complexes, including those involved in secretion (Viollier et al. 2002; Perez-Cheeks et al. 2012), chemotaxis (Thompson et al. 2006; Ringgaard et al. 2011; Alvarado et al. 2017), conjugation (Atmakuri et al. 2007), cell division (Raskin and de Boer 1999; MacCready et al. 2017), and cell motility (Youderian et al. 2003; Kusumoto et al. 2008), as well as protein-based bacterial microcompartments (BMCs), such as the carboxysome (Savage et al. 2010; MacCready et al. 2018). For partitioning plasmids, the ParABS system is the best characterized to date. Mechanistically, ParA proteins dimerize and nonspecifically bind DNA in the presence of ATP (Leonard et al. 2005; Hester and Lutkenhaus 2007; Castaing et al. 2008; Vecchiarelli et al. 2010). Subsequently, ParB, which binds around a centromere-like site on the plasmid, *parS*, displaces ParA from the nucleoid (potentially through ATP hydrolysis stimulation) (Davis and Austin 1988; Funnell 1988; Davis et al. 1992; Bouet and Funnell 1999; Bouet et al. 2000). ParA then recycles its nucleotide and rebinds the nucleoid at a random location. The local formation of ParA depletion zones around individual ParB-bound plasmids results in a global break in ParA symmetry along the nucleoid that ParB-bound plasmids utilize to migrate in a directed and persistent manner toward increased concentrations of ParA on the nucleoid (Adachi et al. 2006; Hatano et al. 2007; HWang et al. 2013; Vecchiarelli et al. 2014; Hu et al. 2017); a recursive mechanism that ensures equidistant plasmid positioning and faithful plasmid inheritance following cell division.

In our recent study, we identified a new self-organizing ParA-type ATPase system, McdAB, which is responsible for equidistantly positioning the carbon-fixing organelles of cyanobacteria, carboxysomes (MacCready et al. 2018). Carboxysomes are essential for photoautotrophic growth of cyanobacteria. Since O_2 competes with CO_2 as a substrate for ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), the encapsulation of RuBisCO and carbonic anhydrase within a selectively permeable protein shell (the carboxysome) is necessary for generating the high CO_2 environment needed to drive internal RuBisCO reactions toward the

Article

Calvin–Benson–Bassham cycle (CO_2 substrate) and away from the wasteful process of photorespiration (O_2 substrate) (fig. 1*A* and *B*) (reviewed in Kerfeld et al. 2018). Comprised thousands of proteins (Sun et al. 2019), carboxysomes were once thought to be completely paracrystalline (Kaneko et al. 2006). However, recent reports have challenged this idea by showing that carboxysome formation involves the process of liquid–liquid phase separation (LLPS) (Oltrogge et al. 2019; Wang et al. 2019); a process that describes the ability of proteins to spontaneously demix into dilute and dense phases that resemble liquid droplets (reviewed in Alberti et al. 2019). Through these mechanisms, carboxysomes contribute to >25% of global carbon-fixation through atmospheric CO_2 assimilation (Rae et al. 2013). Using the model rod-shaped cyanobacterium *Synechococcus elongatus* PCC 7942, we identified a small novel protein, McdB, responsible for emergent oscillatory patterning of ATP-bound McdA on the nucleoid (MacCready et al. 2018). Although McdB had no identifiable sequence similarities with any known ParB-family members, we found that McdB localized to carboxysomes through multiple shell protein interactions and removed McdA from the nucleoid in their vicinity (MacCready et al. 2018); observations that are analogous to ParA patterning following ParB binding to the *parS* site of plasmids. Thus, like plasmids, we showed that carboxysomes utilize a Brownian-ratchet-based mechanism whereby McdB-bound carboxysome motion occurs in a directed and persistent manner toward increased concentrations of McdA on the nucleoid (Vecchiarelli et al. 2014; Hu et al. 2017; MacCready et al. 2018) (fig. 1*C*).

Several important questions remained following our study. First, it was unclear whether the McdAB system was widespread among cyanobacteria, especially given that cyanobacteria are an incredibly diverse and widely distributed phylum of bacteria that display complex morphologies, including: 1) unicellular, 2) baeocystous, 3) filamentous, 4) heterocystous, and 5) ramified (Shih et al. 2013). Cyanobacteria are also taxonomically classified as α or β depending on the form of RuBisCO they encapsulate; β-cyanobacteria encapsulate form 1B RuBisCO in β-carboxysomes and α-cyanobacteria encapsulate form 1 A RuBisCO in α-carboxysomes (Rae et al. 2013). These two types of carboxysomes are structurally distinct, and α-carboxysomes are thought to have been horizontally transferred to cyanobacteria from chemoautotrophic Proteobacteria (Rae et al. 2013). Thus, it is not obvious whether α-cyanobacteria possess a McdAB system and whether it would share similarities to that of the β-cyanobacterial McdAB system. Second, we noted in our previous study that McdA surprisingly lacked the signature lysine residue in the Walker A ATP-binding motif—a critical residue that defines the ParA family of ATPases. Although BLASTp results for McdA identified additional McdA-like sequences in other cyanobacteria that also lacked this lysine residue, the results were few and many were plasmid encoded. Therefore, it was not clear whether the McdAB system was unique to *S. elongatus* or if more than one type of McdAB system evolved in cyanobacteria. Lastly, it was not obvious why reliable results from BLASTp for *S. elongatus* McdB could not be obtained. Interestingly, however, our neighborhood analysis around the carboxysome operon in the distantly related cyanobacterium *Gloeobacter kilaueensis* JS1 identified a ParA-type ATPase that possessed the signature lysine residue absent in *S. elongatus* McdA and a small downstream coding sequence. The protein product of this small gene loaded onto *S. elongatus* carboxysomes, which was surprising because it had no sequence homology to *S. elongatus* McdB (MacCready et al. 2018). The findings suggested a more rigorous gene neighborhood analysis was necessary to identify other McdAB systems.

Here, we performed a neighborhood analysis for McdAB-like sequences encoded near carboxysome components in 537 cyanobacterial genomes (205 α-cyanobacterial and 332 β-cyanobacterial). Our analysis revealed that the McdAB system is widespread among β-cyanobacteria and is surprisingly absent in α-cyanobacteria. Across these β-cyanobacteria, McdAB were found near carboxysome components in ~31% of genomes and near the minor shell components CcmK3 and CcmK4 in ~25% of genomes; suggesting a strong functional association. Our analysis also shows that there are two types of McdAB systems, which we term types 1 and 2. Type 1 systems, like that of *S. elongatus*, consist of a McdA without the signature lysine residue and a McdB with a predicted C-terminal coiled coil. Type 2 systems alternatively, which were found to be the most abundant (>98% of genomes), consist of a McdA ATPase with the signature lysine residue present, and a McdB with a predicted central coiled coil. The low representation of the *S. elongatus* Type 1 McdAB system suggested a possible unique origin. In support of this hypothesis, our cyanobacterial phylogeny strongly suggests that *S. elongatus* is immediately adjacent to α-cyanobacteria, which lack the McdAB system, and shares a common ancestor with cyanobacteria where a McdAB system could not be identified. Lastly, when comparing identified McdB proteins, little to no sequence homology was observed. However, all shared the known hallmarks of proteins capable of LLPS: 1) intrinsic disorder, 2) biased amino acid compositions, 3) low hydrophobicity, and 4) extreme multivalency. We purified representatives of both McdB types and found that all formed phase-separated droplets. Moreover, we found that the LLPS activity of *S. elongatus* McdB was highly influenced by pH; forming droplets in a pH range that correlates to the proposed acidic carboxysome environment and not forming droplets in a pH range that correlates to the basic cyanobacterial cytosol. To our knowledge, this is the first experimental demonstration of LLPS behavior in a ParA-type ATPase partner protein and is an interesting finding given that carboxysome formation has recently been shown to involve LLPS (Oltrogge et al. 2019; Wang et al. 2019). Collectively, these results have broad implications for understanding carboxysome formation, homeostasis, positioning, and function.

Results

Finding Homologs of *S. elongatus* McdAB

To explore whether the McdAB system is widespread among cyanobacteria, we began our analysis by performing BLASTp

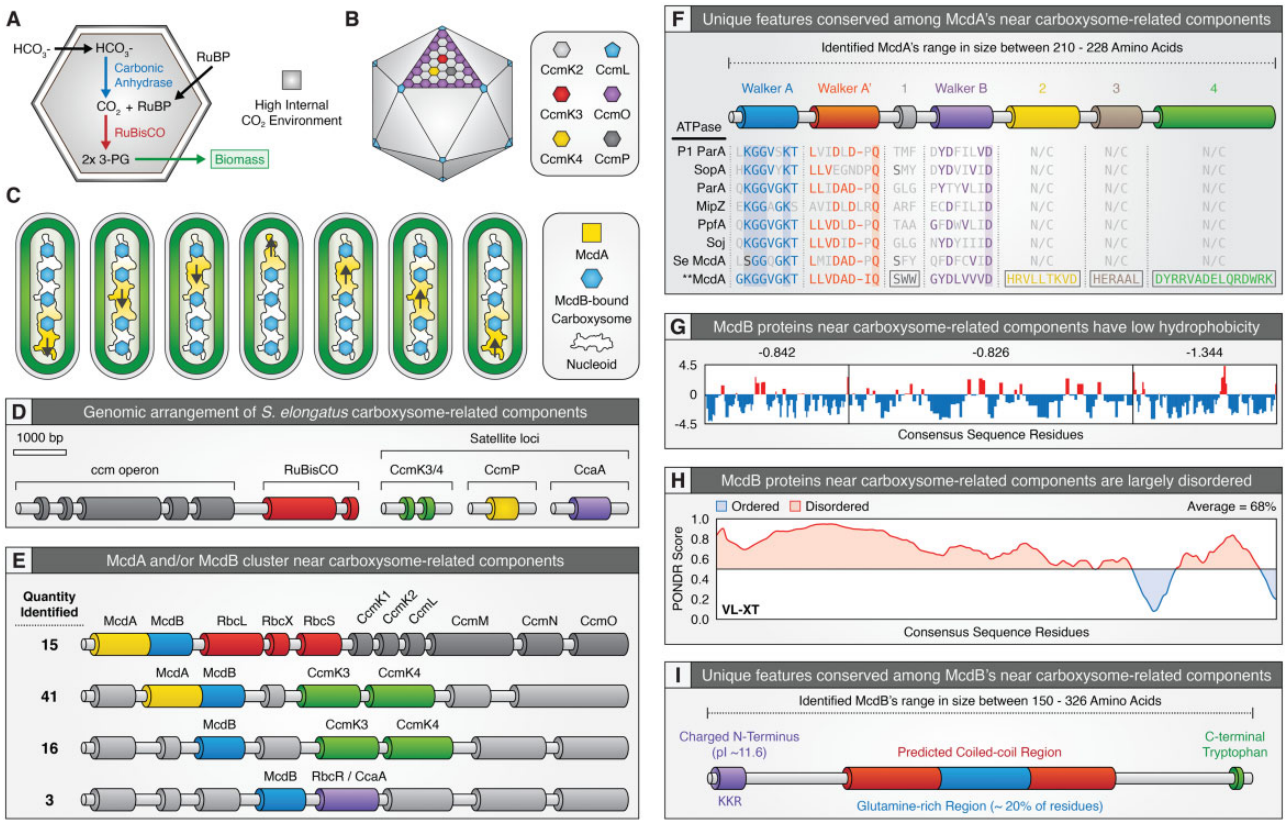

FIG. 1. Candidate McdAB proteins cluster near known carboxysome components and share common unique features. (*A*) Illustration of internal carboxysome enzymatic reactions. (*B*) Illustration of carboxysome protein shell. (*C*) Individual McdB-bound carboxysomes move toward increased concentrations of McdA on the nucleoid that drives equal spacing. (*D*) Representative illustration showing that carboxysome-related genes are found across multiple loci in *Synechococcus elongatus*. (*E*) Representative illustration of the genomic context of McdA and/or McdB near carboxysome-related components. (*F*) Conserved features among McdA proteins found near carboxysome components. Known conserved ParA regions, deviant-Walker A (blue), A′ (red), and B (purple) boxes, are conserved among all classic ParA proteins, *S. elongatus* McdA, and putative McdA proteins identified near carboxysome components (**McdA). Classic ParA ATPase proteins shown: *Escherichia coli* phage P1 ParA (plasmid partitioning—YP_006528), *Escherichia coli* (strain K12) F plasmid SopA (plasmid partitioning—NP_061425), *Caulobacter crescentus* ParA (chromosome segregation—AAB51267), *Caulobacter crescentus* MipZ (chromosome segregation—NP_420968), *Rhodobacter sphaeroides* PpfA (chemotaxis distribution—EGJ21499), and *Bacillus subtilis* Soj (chromosome segregation—NP_391977). Regions conserved among only McdA proteins found near carboxysomes components: Double tryptophan region 1 (gray), and regions 2 (yellow), 3 (brown), and 4 (green). (*G*) Consensus amino acid sequence from identified McdB proteins exhibits a low hydrophobicity. (*H*) McdB proteins are predicted to be intrinsically disordered. (*I*) Conserved features among McdB proteins found near carboxysome components. Charged N-terminal domain (purple), predicted coiled coil (red), glutamine-rich region within coiled coil (blue), and C-terminal tryptophan residue (green).

searches for *S. elongatus* McdA (Synpcc7942_1833) and McdB (Synpcc7942_1834). We previously reported that the deviant Walker A box of *S. elongatus* McdA lacks the signature lysine residue that defines the ParA family of ATPases (**K**GGXXKS/T) (MacCready et al. 2018). The serine substitution at this position in McdA (**S**GGQGKT) may underlie the unusually high ATPase activity of McdA, which displays a maximum specific activity that is roughly two orders of magnitude greater than that of other well-studied ParA systems (Ah-Seng et al. 2009; Vecchiarelli et al. 2010; MacCready et al. 2018). Our BLASTp search results for *S. elongatus* McdA returned only a few McdA-like sequences where the signature lysine residue was replaced with a serine (supplementary fig. S1A, Supplementary Material online). Four of these hits were nearly identical to *S. elongatus* McdA (*Synechococcus elongatus* PCC 6301, *Synechococcus elongatus* PCC 11801, *Synechococcus elongatus* UTEX 3055, and *Synechococcus* sp. UTEX 2973). Recent crystallization of a plasmid-encoded McdA-like protein from the cyanobacterium *Cyanothece* sp. PCC 7424 (PCC7424_5529) showed that a lysine residue in the middle of the protein (K151) functioned analogously to the signature lysine residue of classical ParA-type ATPases found within the Walker A box; K151 was found to contact the oxygen between the β and γ phosphates of ATP and promote formation of a sandwich dimer (Schumacher et al. 2019). Consistent with this finding, our McdA-like BLASTp hits also possessed the K151 residue (supplementary fig. S1A, Supplementary Material online).

BLASTp results for *S. elongatus* McdB were extremely poor. However, when performing a gene neighborhood analysis of the newly identified McdA-like sequences above, a short open reading frame was identified immediately downstream. Although these sequences shared high similarity among themselves, they largely differed from *S. elongatus* McdB

outside the first ~25 amino acids (supplementary fig. S1B, Supplementary Material online). ParB proteins possess a small-charged region at their N-terminus responsible for interacting with its cognate ParA protein and stimulating its ATPase activity (Radnedge et al. 1998; Ravin et al. 2003; Barillà et al. 2007; Ah-Seng et al. 2009). Consistent with this observation, the recent crystallization and analysis of the *Cyanothece* sp. PCC 7424 McdB-like protein (PCC7424_5530) revealed that the N-terminus (AA 1–150) mediated interaction with the *Cyanothece* sp. PCC 7424 McdA-like protein (Schumacher et al. 2019). Given the similarity between *S. elongatus* McdA and the McdA-like sequences, we identified by BLASTp, it is not surprising that the N-terminal region of *S. elongatus* McdB and the McdB-like sequences identified by BLASTp were highly conserved in this region (supplementary fig. S1B, Supplementary Material online). However, outside this N-terminal region, it was not obvious why *S. elongatus* McdB greatly differed from our newly identified McdB-like sequences. The structure of the McdB-like protein from *Cyanothece* sp. PCC 7424 possessed two small central helices followed by a large C-terminal coiled-coil region (Schumacher et al. 2019). Although *S. elongatus* McdB is predicted to also possess a C-terminal coiled coil (MacCready et al. 2018), *S. elongatus* McdB possesses a large glutamine-rich region and several extensions and gaps relative to the McdB-like BLASTp hits (supplementary fig. S1B, Supplementary Material online); potentially due to recognition of different cargos (i.e., carboxysomes vs. plasmids). Indeed, many of these McdAB-like proteins were plasmid encoded and were also the sole ParA-type system on these plasmids. Collectively, given that: 1) several of our McdAB hits were plasmid encoded, suggesting that these proteins might function in plasmid partitioning instead of carboxysome positioning, 2) only a few homologs were identified among the hundreds of available cyanobacterial genomes, 3) the McdB-like protein in *G. kilaueensis* JS1 that we previously showed loaded onto *S. elongatus* carboxysomes (MacCready et al. 2018) followed a ParA-type ATPase that possessed the signature lysine residue and lacked K151, and 4) *G. kilaueensis* JS1 McdAB were encoded within the carboxysome operon, we reasoned that these proteins identified by BLASTp above, as well as the crystallized *Cyanothece* sp. PCC 7424 McdAB-like proteins (Schumacher et al. 2019), were not true McdAB carboxysome positioning systems. Therefore, an alternative more rigorous analysis was needed to identify McdAB homologs across cyanobacteria.

McdAB Cooccur with Carboxysome Components in Many Cyanobacteria

The increased availability of genomic data, in combination with targeted bioinformatic analyses, has resulted in a wealth of data suggesting that the vast majority of BMC-related genes tend to form operons with their respective encapsulated enzymes. For example, the most recent bioinformatic efforts to identify novel BMCs among bacteria resulted in the identification of 23 different types of BMCs across 23 different bacterial phyla (Axen et al. 2014). This study showed that neighborhood analyses are a powerful and reliable tool for the identification of new factors involved with BMC function. Therefore, as BLASTp was an unreliable method to identify new candidate McdA and McdB proteins, we next performed neighborhood analysis for McdAB-like sequences that clustered near carboxysome-related components across 537 cyanobacterial genomes (205 α-cyanobacterial and 332 β-cyanobacterial). We defined candidate McdA sequences as having a deviant Walker A motif with global homology to ParA-type ATPases and candidate McdB sequences as the protein product of the open reading frame immediately following these McdA sequences. We note that additional carboxysome-related genes are often located at distant loci from the main *ccm* (Carbon Concentrating Mechanism) operon (fig. 1D) (Axen et al. 2014; Sommer et al. 2017).

Using these criteria, our analysis identified 75 examples of McdAB clustering near carboxysome components (genomic distances of 2209 $\pm$ 1679 bp or 2 $\pm$ 1.6 genes up- or downstream of the carboxysome component). Of these, McdAB-like sequences clustered near the *ccm* operon in 15 species and near the distant loci of the minor shell proteins CcmK3 and CcmK4 in 41 species (fig. 1E). Moreover, in 16 species, only a McdB-like sequence clustered with CcmK3 and CcmK4, and in three species, a McdB-like sequence was found near either the RuBisCO transcription factor (RbcR) or carbonic anhydrase (CcaA) (fig. 1E). In both these instances, a McdA-like sequence was not identified, suggesting McdB proteins in general may have additional functions independent of its role in positioning carboxysomes via interaction with McdA.

To identify additional McdA and McdB proteins among cyanobacteria that did not cluster near carboxysome components, we needed to establish a new criterion to aid our search. To accomplish this, we generated multiple sequence alignments for the McdA and McdB proteins that clustered near carboxysome components to identify highly conserved regions and/or features among these proteins (supplementary fig. S2A and B, Supplementary Material online). Among the McdA proteins, the first notable features we found were that they could range in size from 210 to 228 amino acids and that each of these proteins possessed the signature lysine residue within the Walker A motif (fig. 1F and supplementary fig. S2A, Supplementary Material online), unlike that of *S. elongatus* McdA which instead possesses a serine residue (MacCready et al. 2018). Secondly, two additional threonine residues followed the last threonine residue of the Walker A motif (supplementary fig. S2A, Supplementary Material online). Lastly, two tryptophan residues adjacent to the Walker A' motif (1—gray cylinder) and three additional regions toward the last half of these proteins (2—yellow cylinder, 3—brown cylinder, and 4—green cylinder) were present among all McdA proteins encoded near carboxysome components, but not among classical ParA proteins or *S. elongatus* McdA (fig. 1F and supplementary fig. S2A, Supplementary Material online).

Unlike these McdA proteins, conservation among McdB proteins near carboxysome components was extremely low; partly due to the surprising observation that McdB proteins ranged in size from 150 to 326 amino acids. However, several

shared features existed that allowed us to establish criteria to define an McdB protein. First, we found that all McdB proteins were largely polar and biased in amino acid composition (fig. 1G). Second, Predictor of Natural Disordered Regions (PONDR) predicted McdB proteins to be highly disordered (average disorder = 68%) (fig. 1H) (Romero et al. 2001), which explains our poor BLASTp results here and prior (MacCready et al. 2018). Lastly, all identified McdB proteins possessed a highly charged N-terminal region, a predicted coiled coil, a glutamine-rich region centrally located within the predicted coiled coils, and a tryptophan residue within the last four residues of each sequence (fig. 1I and supplementary fig. S2B, Supplementary Material online). The bioinformatics analysis suggests all McdB proteins we identified share known hallmarks of phase-separating proteins: largely polar, low amino acid complexity, low hydrophobicity, and intrinsic disorder.

The McdAB System Is Widespread among β-Cyanobacteria

Using our new criteria for McdAB proteins established from McdAB proteins that cluster near carboxysome components, we performed an additional manual search within α- and β-cyanobacterial genomes where McdAB proteins were not previously identified. This approach was necessary given that cyanobacteria possess enormously diverse genomic architectures and relatively poor operon structure in comparison to other well-studied bacteria (Beck et al. 2018). Not a single McdAB system that fit within our criteria was identified among the 205 α-cyanobacterial genomes we analyzed. Intriguingly, many α-cyanobacteria did not possess a single homolog of a ParA-type ATPase family member in general. However, in β-cyanobacteria, we identified several McdA and McdB sequences that fit within our criteria. In total, we found 248 McdA (∼75% of β-cyanobacterial genomes) and 285 McdB sequences (∼86% β-cyanobacterial genomes) from 332 β-cyanobacterial genomes analyzed (fig. 2A and supplementary table S1, Supplementary Material online). We found that ∼14% β-cyanobacterial genomes lacked the McdAB system. However, we note that 16% of β-cyanobacterial genomes without candidate McdA, 8% of β-cyanobacterial genomes without McdB, and 9% of β-cyanobacterial genomes without McdA and McdB were not fully assembled genomes (still in scaffolds or contigs). Collectively, we found that McdA proteins ranged in lengths from 202 to 253 amino acids, while the lengths of McdB proteins varied far greater, ranging in from 132 to 394 amino acids. In 199 β-cyanobacterial genomes, McdAB were found next to each other genomically (∼82% of β-cyanobacterial genomes with McdAB) (fig. 2A). Moreover, McdA and/or McdB clustered near carboxysome-related components in 75 genomes (∼31% of β-cyanobacterial genomes with McdAB) and clustered near the distant locus for the minor shell components CcmK3 and CcmK4 in 60 genomes (∼25% of β-cyanobacterial genomes with McdAB). This finding suggests a strong functional association between McdB and CcmK3/CcmK4, which is consistent with our previous bacterial two-hybrid results showing a strong interaction between McdB and these two shell components of the carboxysome (MacCready et al. 2018). Lastly, McdAB sequences were found not only among all major taxonomic orders of β-cyanobacteria (fig. 2B) but were also found in genomes across all five major morphologies (fig. 2C).

Two Distinct McdAB Systems Exist in β-Cyanobacteria

While conducting this bioinformatic analysis, we found that the vast majority (∼98%) of newly identified McdAB proteins was quite distinct from *S. elongatus* McdAB and only four additional cyanobacterial species, which are closely related to *S. elongatus*, had a similar McdAB system to that of *S. elongatus*. Therefore, we designated the two McdAB systems as Type 1, those similar to *S. elongatus* McdAB, and Type 2, those similar to the newly identified McdAB's in this study (fig. 2D and E). Several features distinguish McdA as Type 1 or Type 2. For example, Type 1 McdA proteins lack the signature lysine residue within the Walker A motif, whereas Type 2 McdA's possess this lysine residue like classical ParA family members. The double tryptophan found in all Type 2 McdA proteins (*) is instead a phenylalanine and a tyrosine in Type 1 McdAs (fig. 2D). Additionally, a small ∼7 amino acid insertion (**) immediately following the double tryptophan as well as a downstream phenylalanine are highly conserved in Type 2 McdA's, but absent in Type 1 (fig. 2D). Most notably, Type 1 McdA's have a long extension between the Walker A' and Walker B motifs that is not present in Type 2 McdA's (fig. 2D). Lastly, the Walker B motif of Type 1 McdA's has a conserved phenylalanine and cysteine that are instead small hydrophobic residues in Type 2 McdA's.

We also found the two types of McdB significantly differed. For example, Type 1 McdB's have a predicted coiled coil at the C-terminus, while all Type 2 McdB's have a predicted coiled coil near the middle of the protein (fig. 2E). Moreover, the charged N-terminal region appeared inverted among Type 1 and Type 2 McdB's, where negatively charged residues proceeded positively charged residues in Type 1 McdB's and positively charged residues proceeded negatively charged residues in Type 2 (fig. 2E). Interestingly, although both types have a central glutamine-rich region, Type 1 McdB's are almost 10% more enriched for glutamine than Type 2 McdB's (fig. 2E). Despite these differences, we found that both types possessed an invariant C-terminal tryptophan (fig. 2E). Taken together, these features strongly suggest that two distinct McdAB systems are present among cyanobacteria and that the Type 2 system identified in this study is the most widely conserved.

A Unique Origin for Type 1 McdAB Systems?

Since we found that α-cyanobacteria lack McdAB and that Type 1 systems were only present in ∼2% of cyanobacteria studied, we next sought to better understand the phylogenetic placement of Type 1 and Type 2 McdAB systems and determine whether the Type 1 McdAB system could have evolved independent of Type 2 systems. To explore this, we generated a Maximum-Likelihood tree inferred using a concatenation of the proteins DnaG, RplA, RplB, RplC, RplD, and

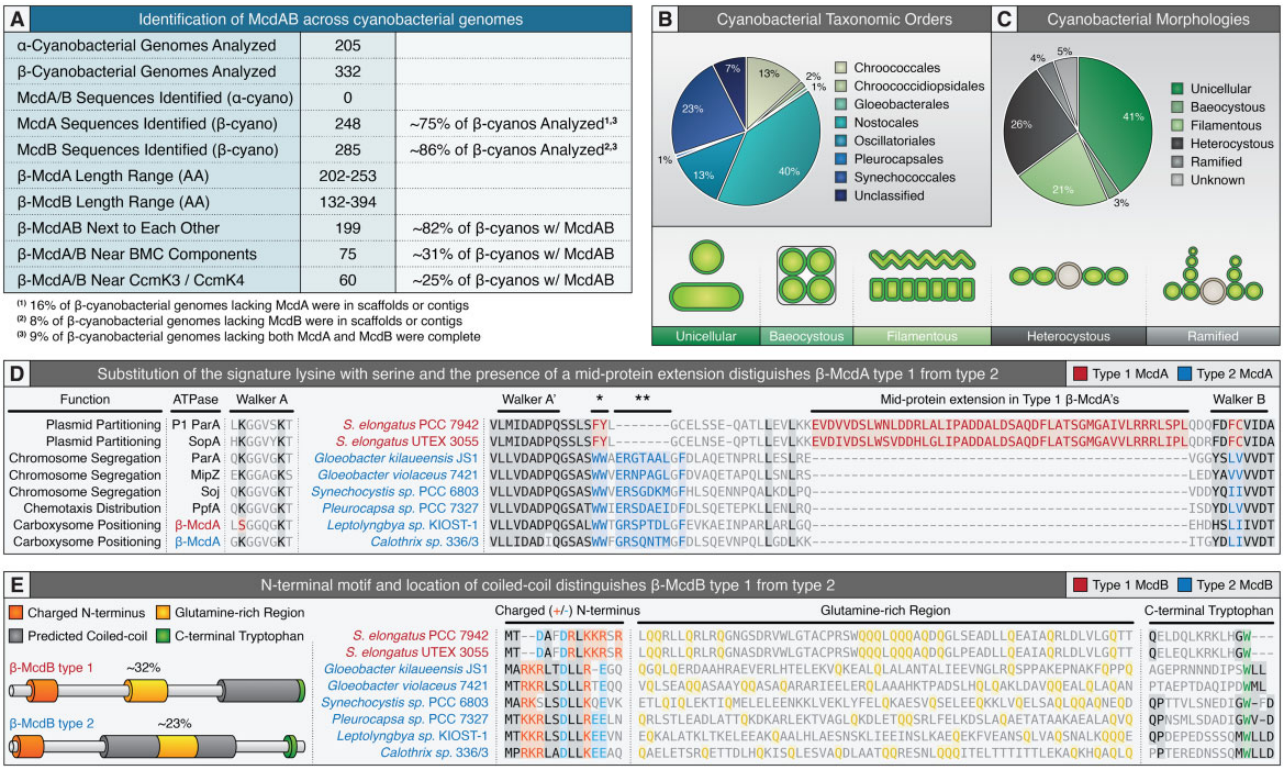

FIG. 2. Two distinct McdAB systems exist in β-cyanobacteria. (*A*) Table highlighting the prevalence of certain sequence features for all McdAB proteins identified among cyanobacteria. (*B*) McdAB are widely distributed among cyanobacterial taxonomic orders. (*C*) McdAB are found in all five major morphologies of cyanobacteria. General illustration of cyanobacterial morphologies below. (*D*) Type 1 McdA proteins (red) are distinct from Type 2 McdA proteins (blue). Left: Type 1 McdA proteins (red) possess a serine instead of lysine in the Walker A box. Type 2 McdA proteins (blue) possess the signature lysine. Middle: Areas of conservation are shaded black. Conserved regions unique to Type 1 McdA proteins shaded red. Conserved regions unique to Type 2 McdA proteins shaded blue. The Walker A′ box is conserved among both McdA types. Type 2 McdA proteins have a highly conserved double tryptophan region (*) not found in Type 1. Type 2 McdA proteins have a small 7 amino acid insertion (**) and highly conserved phenylalanine, glutamic acid, and proline residues following the double tryptophan region. Type 1 McdA proteins have a large internal extension not found in Type 2. Right: Walker B box is generally conserved, but Type 1 McdA proteins possess phenylalanine and cysteine residues that are instead small polar residues in Type 2 McdA proteins. McdA sequences shown: *Synechococcus elongatus* PCC 7942 (Synpcc7942_1833), *Synechococcus elongatus* UTEX 3055 (Unannotated), *Gloeobacter kilaueensis* JS1 (GKIL_0670), *Gloeobacter violaceus* 7421 (glr2463), *Synechocystis* sp. PCC 6803 (MYO_127120), *Pleurocapsa* sp. PCC 7327 (Ple7327_2492), *Leptolyngbya* sp. KIOST-1 (WP_081972678), and *Calothrix* sp. 336/3 (AKG24853). (*E*) Type 1 McdB proteins (red) are distinct from Type 2 McdB proteins (blue). Left: Type 1 McdB proteins have a charged N-terminal domain (orange), central glutamine-rich region (yellow), C-terminal coiled coil (gray), and terminal tryptophan residue (green). Type 2 McdB proteins have a charged N-terminal domain (orange), central coiled coil (gray), glutamine-rich regions within the coiled coil (yellow), and a C-terminal tryptophan within the last four amino acids. Middle: The N-terminal charged (positive charge—orange, negative charge—blue) domain is inverted between Type 1 and Type 2 McdB proteins. Middle: Glutamine-rich regions (yellow). Right: All McdB proteins have a tryptophan residue within the last four amino acids (green). McdB sequences shown: *Synechococcus elongatus* PCC 7942 (Synpcc7942_1834), *Synechococcus elongatus* UTEX 3055 (Unannotated), *Gloeobacter kilaueensis* JS1 (GKIL_0671), *Gloeobacter violaceus* 7421 (glr2464), *Synechocystis* sp. PCC 6803 (MYO_127130), *Pleurocapsa* sp. PCC 7327 (Ple7327_2493), *Leptolyngbya* sp. KIOST-1 (WP_035984653), and *Calothrix* sp. 336/3 (Unannotated).

RplE, which have recently been shown to be good-markers for cyanobacterial phylogenetic reconstruction (Hirose et al. 2019). Notably, we found that the Type 2 McdAB system was widespread among cyanobacteria including our outgroup taxonomic order Gloeobacterales, considered the "most primitive" order among living cyanobacteria due to a lack of thylakoid membranes (Rippka et al. 1974; Guglielmi et al. 1981; Nakamura et al. 2003). This was an important finding that suggested the Type 2 McdAB system was present at the earliest known point of cyanobacterial evolution. Alternatively, the Type 1 McdAB system was only present in cyanobacterial species adjacent to α-cyanobacteria, which we found to lack the McdAB system, and shared a common ancestor with four β-cyanobacterial species where we could not identify an McdAB system of either type (fig. 3). This was a surprising result that suggested a more recent origin for the Type 1 McdAB system. Our phylogenic inference also suggested multiple independent losses of the McdAB system with no obvious shared characteristics among species (i.e., different taxonomic orders and morphologies) (fig. 3). Together, these results suggest that the Type 2 McdAB system is more ancestral and that Type 1 evolved independently of the Type 2 McdAB system, possibly via horizontal gene transfer.

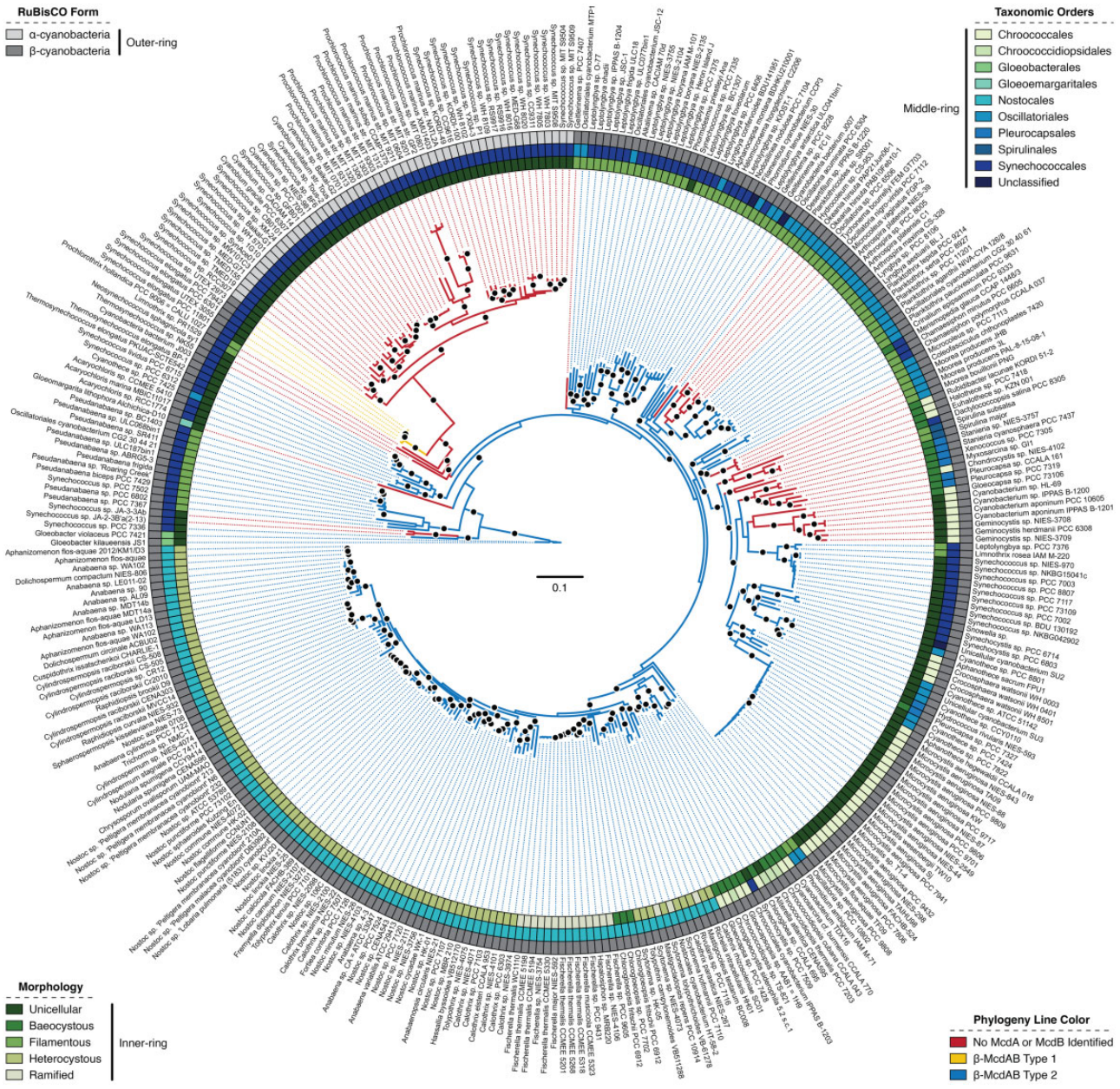

FIG. 3. A possible unique origin for the Type 1 McdAB system. Inferred cyanobacterial phylogeny of genomes analyzed. Outer ring: cyanobacterial RuBisCO type. Middle ring: cyanobacterial taxonomic order. Inner ring: cyanobacterial morphology. Line color: Type 1 McdAB systems (yellow), Type 2 McdAB system (blue), and no identified McdAB system (red). Black dot represents >70% support (500 replicates).

McdB Proteins Possess the Hallmarks for Phase Separation

Why McdB proteins were so highly diverged in primary sequence was unclear. For example, while Type 2 McdA's have high mean amino acid similarity (~64%), we found that the mean amino acid similarity among Type 2 McdB's was extremely low (~20%) (fig. 4A and B). Partially contributing to this diversity among McdB proteins, we found that the lengths of the central coiled coils and the N- and C-terminal extensions from the coiled coil greatly varied in length (fig. 4C–E). N-terminal extensions were 60 ± 23 amino acids, central coiled coils were 93 ± 25 amino acids, and the C-terminal extensions were 56 ± 16 amino acids in length (fig. 4C–E). ProtScale prediction (Scale: Kyte and Doolittle 1982) determined that both Type 1 and Type 2 McdB's were largely hydrophilic, and that Type 1 McdB's possessed a small hydrophobic patch toward the middle of the protein (fig. 4F). Consistent with this finding, all three Type 2 McdB domains were biased toward hydrophilic amino acids and had obvious repetitive sequences that were rarely similar from one McdB protein to the next (fig. 4G).

Consistent with the low hydrophobicity of McdB proteins, PONDR predicted that McdB proteins were largely disordered (fig. 4H). Indeed, we found that the mean disorder of McdB proteins was 64%; PONDR predicted that some McdB proteins were as high as 96% disordered (fig. 4H). Interestingly, while the consensus sequence for Type 1 McdB's was predicted to be largely ordered in a central region that partially

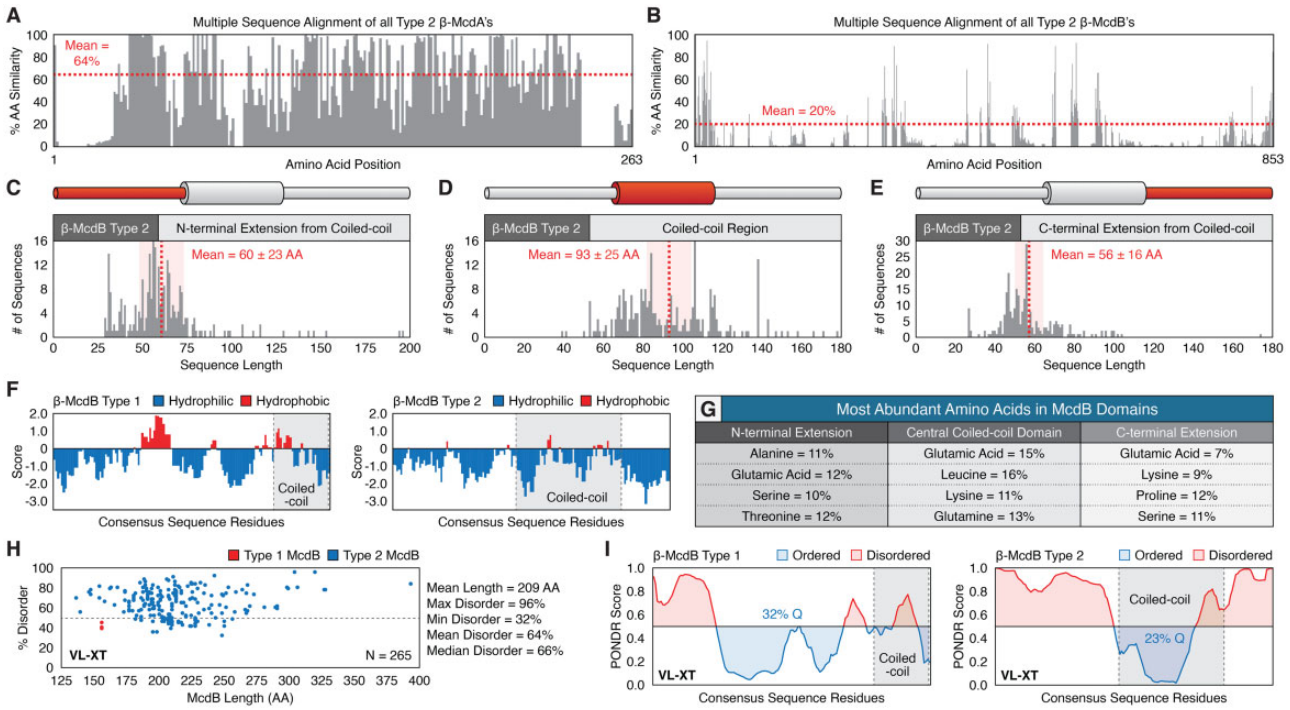

FIG. 4. Two distinct McdAB systems exist in β-cyanobacteria. (*A*) Plot of amino acid similarity from multiple sequence alignment of Type 2 McdA proteins (*n*=XX sequences). (*B*) Plot of amino acid similarity from multiple sequence alignment of Type 2 McdB proteins (*n*=XX sequences). (*C*) Plot of Type 2 McdB N-terminal extension lengths, (*D*) coiled-coil lengths, and (*E*) C-terminal extension lengths. SD shaded red behind the mean. (*F*) Comparison of hydrophobicity between consensus sequences of Type 1 (left) and Type 2 (right) McdB proteins. (*G*) Table quantifying biased amino acid compositions among McdB protein domains. (*H*) PONDR disorder scatter plot for all Type 1 (red) and Type 2 (blue) McdB proteins. (*I*) PONDER disorder plot between consensus sequences of Type 1 and Type 2 McdB proteins.

corresponded to a hydrophobic patch and that the bulk of disorder was positioned toward the N-terminal region of the proteins (fig. 4*F* and *I*), the consensus sequence for Type 2 McdB's was predicted to possess much greater disorder in the N- and C-terminal extensions from the central coiled coils and predicted to be ordered in a region that corresponded to the coiled coils (fig. 4*I*). Collectively, these features were largely responsible for the diversity among McdB proteins.

McdB Undergoes LLPS In Vitro

The shared hallmarks of McdB proteins include: 1) large regions of low complexity that greatly vary in length, 2) intrinsically disordered regions, 3) repetitive and biased amino acid compositions, 4) low hydrophobicity, and 5) extreme multivalency. All are features that are characteristic of proteins shown to undergo LLPS (Kato et al. 2012; Oldfield and Dunker 2014; Elbaum-Garfinkle et al. 2015; Lin et al. 2015; Molliex et al. 2015; Nott et al. 2015; Varadi et al. 2015; Jain et al. 2016). LLPS refers to the ability of an otherwise homogeneous solution of macromolecules (e.g., proteins or nucleic acids) to spontaneously demix into a dilute phase and dense phase that resembles water droplets (fig. 5*A*) (reviewed in Alberti et al. 2019). The two liquid-like phases coexist and can in some cases undergo further reversible phase transitions to form gels and solids depending on solution conditions (i.e., macromolecule concentration, pH, salt type and concentration, and temperature); some transitions are irreversible under physiological conditions, such as amyloid-like fibers (Alberti et al. 2019). This phenomenon has received increased attention in cell biology due to the discovery that this process accounts for the compartmentalization of biological functions through the formation of membraneless organelles, also recently termed biomolecular condensates (Shin et al. 2017).

Given the general features of the McdB proteins, we identified in this study (fig. 4), we next wanted to test the hypothesis that *S. elongatus* McdB could undergo phase separation. We expressed and purified *S. elongatus* McdB fused at its N-terminus with a 6×-histidine tag for affinity purification and a SUMO tag to form His–SUMO–McdB. A SUMO-tag has been shown to inhibit the LLPS activity of proteins, thus simplifying the purification protocol (Schuster et al. 2018). In addition, cleavage of the SUMO-tag by the addition of the Ulp1 protease allows for precise, time-controlled induction of LLPS. Under physiologically relevant pH and salt concentration (pH 7.5, 150 mM KCl), native *S. elongatus* McdB displayed phase-separation upon the addition of Ulp1 (fig. 5*B*).

The local environment of the carboxysome in *S. elongatus* cells has been suggested to be more acidic relative to the more basic cytosol because of bicarbonate accumulation prior to its conversion to CO_2 by carbonic anhydrase (Mangan et al. 2016). Given that McdB strongly colocalizes with carboxysomes in vivo (MacCready et al. 2018), we assayed the LLPS activity of McdB across a broad pH range (5.5–8.5) (fig. 5*B*). We found that McdB (30 μM) readily formed droplets following cleavage of the His–SUMO tag

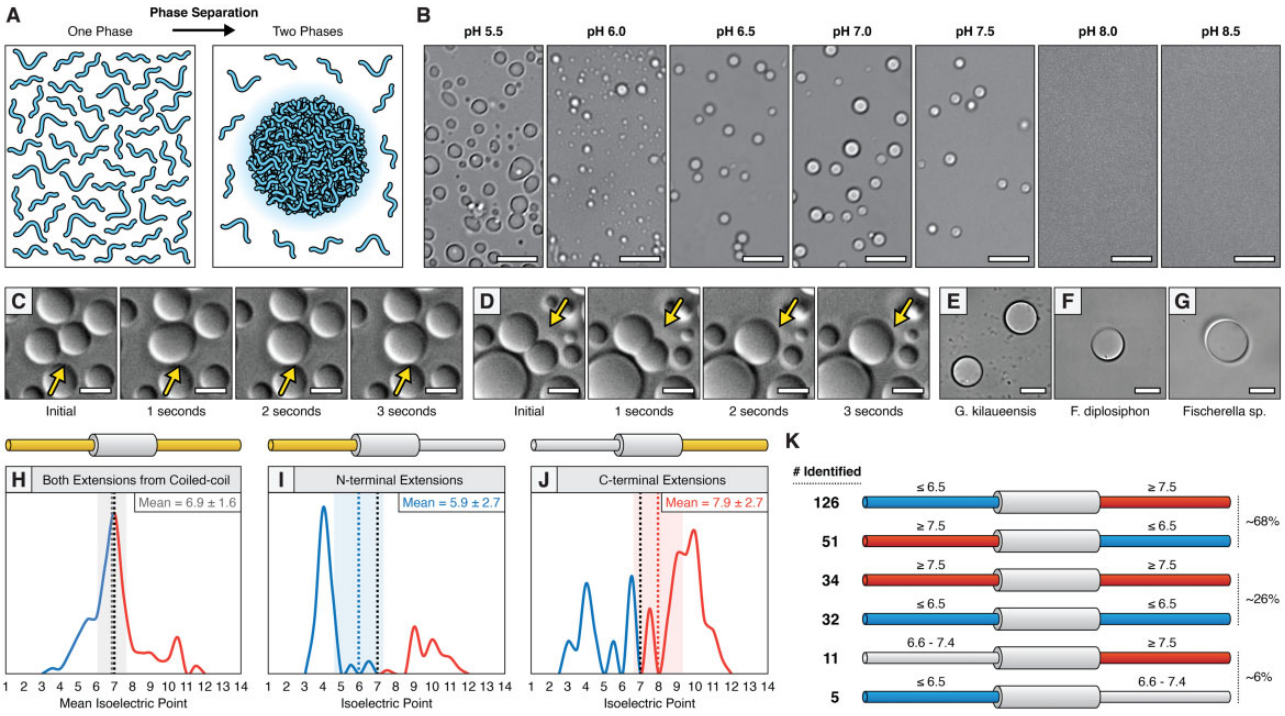

FIG. 5. *Synechococcus elongatus* McdB undergoes liquid–liquid phase separation. (*A*) Cartoon illustration of protein liquid–liquid phase separation from one phase to two phases. (*B*) Microscopy images of *S. elongatus* McdB droplets under varying pH. Scale bar = 10 μm. (*C* and *D*) McdB droplet fusion events (yellow arrows). Scale bar = 5 μm. (*E*) *Gloeobacter kilaueensis* JS1 McdB droplets at pH 7.0. Scale bar = 10 μm. (*F*) *Fremyella diplosiphon* NIES-3275 McdB droplets at pH 7.0. Scale bar = 10 μm. (*G*) *Fischerella* sp. PCC 9431 McdB droplets at pH 7.0. Scale bar = 10 μm. (*H*) Plot of the isoelectric point for both N- and C-terminal extensions of Type 2 McdB proteins. (*I*) Plot of the isoelectric point for the N-terminal extensions of Type 2 McdB proteins. (*J*) Plot of the isoelectric point for the C-terminal extensions of Type 2 McdB proteins. (*K*) Illustration of the number of Type 2 McdB proteins identified that have patterned charge distributions. Isoelectric point range listed above each extension.

across a pH range of 5.5–7.5 (fig. 5*B*). Between pH 5.5 and 6.5, droplets were less spherical in shape and did not readily fuse, which is suggestive of a gel-state (Alberti et al. 2019). However, between pH 6.5–7.5, McdB droplets displayed liquid-like behaviors such as the fusion of two adjacent droplets into one (fig. 5*B*). At or above pH 8, McdB did not undergo LLPS (fig. 5*B*). Even with significantly higher concentrations of McdB (167 μM), McdB still displayed liquid-like behavior with frequent fusion events (fig. 5*C* and *D* and supplementary videos 1 and 2, Supplementary Material online). McdB formed droplets at concentrations as low as 1 μM, but the droplets were very small and difficult to image due to their rapid diffusion in the observation well. Complete cleavage of the His–SUMO tag was verified across the assayed pH range (supplementary fig. S3, Supplementary Material online). The data suggest that the acidic nature of the carboxysome, relative to the cytosol of *S. elongatus* at pH 8 (Mangan et al. 2016), may facilitate local phase-separation of McdB in the vicinity of carboxysomes.

Both McdB types share the same hallmark features for LLPS at the amino acid sequence level. However, Type 2 McdBs differ significantly in coiled-coil location, charge distribution, and amino acid sequence length. Therefore, we next assayed whether Type 2 McdB proteins also displayed LLPS activity, in spite of these differences. As with *S. elongatus* McdB, we expressed and purified three Type 2 McdB proteins fused at their N-termini with a 6×-histidine and SUMO tag. These three Type 2 McdB proteins were strategically chosen for their broad phylogenetic distribution and presence in morphologically diverse cyanobacteria—*Gloeobacter kilaueensis* JS1, *Fremyella diplosiphon* NIES-3275, and *Fischerella* sp. PCC 9431 McdB proteins (fig. 3). For all three Type 2 McdB proteins tested, we observed phase separation into droplets in vitro (fig. 5*E*–*G*). Therefore, we conclude that LLPS activity is likely a universal property in both McdB types identified here. To our knowledge, this is the first direct demonstration of LLPS behavior of a ParA partner protein.

Charge Distribution May Contribute to McdB Phase Separation

Multivalency (charge distribution in the primary amino acid sequence) is an important feature contributing to LLPS (Li et al. 2012; Harmon et al. 2017). For example, LLPS among the most well-studied systems involves interaction between positively charged residues within an intrinsically disordered region of proteins and the negatively charged backbone of DNA or RNA (Alberti et al. 2019). However, we previously showed that McdB does not interact with DNA, but instead interacts with the carboxysome shell proteins CcmK2, CcmK3, CcmK4, CcmL, and CcmO (MacCready et al. 2018). We asked where the multivalency within our system could arise. Although shell proteins have surface accessible acidic and basic residue patches that McdB could interact with, shell proteins are not intrinsically disordered nor are the acidic and basic regions of

all shell proteins that McdB interacts with similar (Sommer et al. 2017). Therefore, given that *S. elongatus* McdB could robustly phase separate on its own, without the addition of any interacting partner (i.e., a carboxysome shell component), we next set out to determine whether Type 1 and Type 2 McdB proteins were intrinsically multivalent.

Analysis of the single N-terminal extension from the C-terminal coiled coil of *S. elongatus* Type 1 McdB revealed that the isoelectric point of roughly the first half of the extension was ~8.5 (amino acids 1–50) and the second half was ~4.3 (amino acids 51–108). For Type 2 McdB proteins that have N- and C-terminal extensions from a central coiled coil, the average isoelectric point of both extensions combined was essentially neutral (6.9 ± 1.6) (fig. 5*H*). Intriguingly, however, we found that the average isoelectric point of the vast majority of N- and C-terminal extensions was either negatively- or positively charged (fig. 5*I* and *J*). Moreover, when looking at Type 2 McdB proteins individually, we found that the vast majority (~68%) had extremely inverted charges distributed between their N- and C-terminal extensions; ~48% had a negatively charged N-terminus and positively charged C-terminus, and ~20% had a positively charged N-terminus and negatively charged C-terminus (fig. 5*K* and supplementary fig. S4, Supplementary Material online). Roughly 26% of McdB proteins had both extensions share a similar charge with 13% having both positively charged extensions, and ~13% having both negatively charged extensions (fig. 5*K* and supplementary fig. S4, Supplementary Material online). Lastly, only ~6% of McdB proteins had one charged extension and one neutral extension; ~4% had a neutral N-terminus and positively charged C-terminus and ~2% had a negatively charged N-terminus and neutral C-terminus (fig. 5*K* and supplementary fig. S4, Supplementary Material online). Together, these results suggest that McdB is a polyampholyte with biphasic charge distributions within intrinsically disordered regions that may contribute to its LLPS activity.

Discussion

McdAB Systems Are Widespread among β-Cyanobacteria

Carboxysomes are essential protein-based organelles in cyanobacteria. To ensure that each daughter cell receives an optimum quantity of carboxysomes following cell division, the McdAB system equidistantly positions each carboxysome relative to one another throughout the cell (fig. 1*C*). In our previous study, we were unable to determine how widespread the McdAB system was among cyanobacteria. Although we were able to identify one McdAB system outside of *S. elongatus*, in *Gloeobacter kilaueensis* JS1, and show that its McdB protein was capable of loading onto carboxysomes in *S. elongatus*, it was not obvious why BLASTp results for *S. elongatus* McdAB were so poor and why *G. kilaueensis* JS1 McdAB were so different than those from *S. elongatus* (22.5% and 18.4% pairwise identity to *S. elongatus* McdAB) (MacCready et al. 2018).

As McdAB were situated near the carboxysome operon in *G. kilaueensis* JS1, we reasoned that neighborhood analysis was a better method to identify McdAB proteins in other cyanobacteria. Indeed, we were able to identify 56 McdA and 75 McdB sequences that clustered near carboxysome components throughout cyanobacteria (fig. 1*E* and supplementary table S1, Supplementary Material online). Given this much larger sample size of McdAB proteins, we were able to identify highly conserved regions and features that permitted further identification of McdAB sequences that did not cluster near carboxysome components in other cyanobacteria. In total, we identified 248 McdA and 285 McdB sequences, showing that the McdAB system is widespread among β-cyanobacteria (fig. 2*A*).

These results have broader implications for understanding carboxysome positioning in other species. For example, cyanobacteria can display a wide range of morphologies from unicellular to specialized multicellular (fig. 2*C*). Although carboxysomes are linearly spaced or hexagonally packed along the long axis of rod-shaped *S. elongatus* cells, our prior modeling suggested that McdAB positioning is influenced by cellular geometry, but still operates within spherical cyanobacterial cells to optimally space carboxysomes from one another (MacCready et al. 2018). Likewise, many heterocystous and ramified cyanobacteria, while appearing filamentous, are a chain of connected smaller cells that individually display more rounded morphologies. Transmission electron micrographs from heterocystous cyanobacteria reveal that the McdAB system likely hexagonally packs carboxysomes (Montgomery 2015). Therefore, given the ubiquity of McdAB across these morphologies (fig. 2*C*), an understanding of how this system behaves within these unique cellular geometries is of profound interest.

Lastly, we found that a large clade of cyanobacteria on the right side of our phylogenic tree appeared to lack the McdAB system (fig. 3). Many of these species, including *Myxosarcina* sp. GI1, *Xenococcus* sp. PCC 7305, *Stanieria* sp. NIES-3757, *Stanieria cyanosphaera* PCC 7437, *Pleurocapsa* sp. CCALA 161, and *Pleurocapsa* sp. PCC 7319, display baeocystous morphologies and some species, including *Spirulina major* and *Spirulina subsalsa*, while filamentous, have extremely spiralized morphologies. One possible reason for the absence of the McdAB system in these cellular morphologies is that their chromosome organization, cell growth, and division strategies may be incompatible.

Two Types of McdAB Systems Exist in β-Cyanobacteria

Our analysis has identified two distinct McdAB systems in β-cyanobacteria. Type 2 systems are by far the most represented among cyanobacteria (>98% of species). This finding in particular explains why BLASTp results for the Type 1 system of *S. elongatus* McdAB were so poor. There are several features that distinguish Type 1 and Type 2 McdAB systems. For example, Type 2 McdA proteins possess the "signature lysine" residue in the Walker A box ATP-binding motif; a feature that defines this family of ATPases as "ParA-like." This result was intriguing given that Type 1 McdA proteins, like that of *S. elongatus*, lack this signature lysine in the Walker A box. Instead, Type 1 McdA proteins likely use a lysine residue

(K151) distantly located on the C-terminal half of the protein to fulfill this role (Schumacher et al. 2019) (fig. 2*D*). In addition, exclusive to Type 1 McdA proteins is a large midprotein extension adjacent to the Walker B box. Overall, it is the Type 2 McdA protein that bears the closest similarity to ParA-family ATPases.

The differences in amino acid sequence in and around the ATP-binding motifs suggest these two McdA types may also have differences in ATPase activity. Indeed, we previously found that the Type 1 McdA of *S. elongatus* has voracious ATPase activity and a relatively low ATPase stimulation by McdB (MacCready et al. 2018). This is in stark contrast to ParA family members involved in bacterial DNA segregation, which have feeble intrinsic ATPase activities, but are strongly stimulated by their cognate ParB protein. Therefore, it will be interesting to compare the ATPase activities of the two McdA types, as well as the stimulatory activities of their respective McdB proteins. How Type 1 and Type 2 systems differ mechanistically, and whether these differences correlate to known factors influencing carboxysome positioning such as carboxysome size and quantity (Gonzalez-Esquer et al. 2016; MacCready et al. 2018), genome copy number and volume (Griese et al. 2011), redox state of cells (Sun et al. 2016, 2019), and McdAB stoichiometry (MacCready et al. 2018), as well as unknown factors including McdB LLPS behavior and/or intracellular salt/pH conditions (see below), is important for understanding the diversity and evolution of the McdAB system among cyanobacteria.

Although McdA proteins differed in certain regions, they were largely conserved compared with McdB proteins that displayed extreme diversity. The lack in conservation is presumably a result of relaxed selection, as any two closely related cyanobacterial species had McdB proteins that poorly aligned (supplementary fig. S2*B*, Supplementary Material online). Despite this lack in conservation, we were able to identify several shared features among McdB proteins. All possessed a highly charged N-terminus, a glutamine-rich region toward the center of the protein, a coiled-coil domain, and an invariant tryptophan residue within the last four amino acids of the protein (fig. 2*E*). For several ParA family ATPases, the N-terminus of the partner protein is necessary and sufficient for stimulation of ParA ATPase activity (Radnedge et al. 1998; Ravin et al. 2003; Barillà et al. 2007; Ah-Seng et al. 2009). Where tested, a critical N-terminal arginine or lysine residue stimulates the ATPase activity of the cognate ParA. The charged N-terminal domains of Type 1 versus Type 2 McdB proteins appeared inverted (location of positively- and negatively charged residues) (fig. 2*E*). This finding suggests a possible difference in how Type 1 and Type 2 McdB proteins interact with their McdA partner, potentially through this N-terminal region. The central glutamine-rich region likely plays a role in LLPS activity of McdB proteins (see below). A striking difference between the two McdB types was the placement of the coiled coil. Type 1 McdB proteins have a coiled coil positioned at their C-terminus, whereas the coiled coil of Type 2 McdB proteins was centrally located (supplementary fig. S4, Supplementary Material online). Coiled coils provide an interface for stable oligomerization into defined higher order species (Apostolovic et al. 2010). Consistent with our sequence-based predictions, a recently solved structure of the coiled-coil domain in a Type 1 McdB-like protein showed stable dimerization in an antiparallel fashion (Schumacher et al. 2019). As for the invariant tryptophan at the C-terminus of all McdBs, it is intriguing that many proteins involved in the assembly of viral- or phage-capsids also encode for a tryptophan residue at their C-terminus (Deeb 1973; Tsuboi et al. 2003; Komla-Soukha and Sureau 2006; Marintcheva et al. 2006). Given the capsid-like icosahedral structure of the carboxysome, it is attractive to speculate that the C-terminal tryptophan of McdB is involved in its association with the carboxysome shell.

The functions and features we have shown thus far for Type 2 McdAB systems strongly suggest that carboxysomes are positioned by Type 2 systems by a very similar mechanism to that of Type 1 systems. Type 2 McdAB genes frequently neighbor carboxysome-related components, both McdB types have similar domain architectures and undergo LLPS, and we previously showed that a Type 2 McdB from *G. kilaueensis* JS1 is capable of loading onto *S. elongatus* carboxysomes (MacCready et al. 2018). It was also recently shown that carboxysomes in *Fremyella diplosiphon*, which possesses a Type 2 McdAB system, are positioned similar to that shown for *S. elongatus* with a Type 1 McdAB system (Lechno-Yossef et al. 2019). However, it remains speculative at this point whether Type 2 McdAB systems are indeed responsible for positioning carboxysomes in vivo. A full characterization of a Type 2 McdAB system in its native host will be required to confirm their role in positioning carboxysomes, for elucidating potential biochemical differences among the two systems, and for studying how different cyanobacterial morphologies influence McdAB dynamics.

The McdAB System Is Not Present in α-Cyanobacteria

One of our most striking findings was that the McdAB system was completely absent in α-cyanobacteria. Although there are several differences between α- and β-carboxysomes, we expected this difference would largely be reflected in the putative carboxysome interaction domains of McdB proteins. We did not anticipate a complete absence of the system. Even more surprising, we found that not just McdA, but the entire ParA family of ATPases were significantly underrepresented among α-cyanobacteria. α-Cyanobacteria lack plasmids and have small genomes (most *Prochlorococcus* genomes are smaller than 2 Mb) (Scanlan et al. 2009). Therefore, one possible explanation for the lack of ParA-type ATPases is that many canonical partitioning systems might not be able to function in α-cyanobacteria. In addition, the absence of plasmids in α-cyanobacteria limits one of the sources of horizontal gene transfer for inheriting the McdAB system.

Alternative positioning mechanisms may exist for α-carboxysomes in α-cyanobacteria. For example, α-carboxysomes are known to tightly interact with polyphosphate bodies (Iancu et al. 2010). If polyphosphate bodies are strategically positioned in α-cyanobacteria, and α-carboxysomes tightly interact with these structures, this interaction would provide a "pilot-fish" mechanism by which both are equidistantly

positioned throughout cells prior to cell division; making the McdAB system unnecessary. Another possibility is that α-carboxysomes are actively positioned by a polymer-based system, such as the actin-like ATPase MamK that mediates magnetosome positioning (Toro-Nahuelpan et al. 2016) or the tubulin-like GTPase TubZ that positions viral DNA (Oliva et al. 2012).

As α-carboxysomes are proposed to have originated in Proteobacteria and were horizontally transferred into cyanobacteria (Marin et al. 2007; Rae et al. 2013), it was surprising that the McdAB system was absent in α-cyanobacteria largely because many carboxysome operons of Proteobacteria encode a ParA-type ATPase followed by a small coding sequence. Whether this ParA-type ATPase and the small downstream coding sequence constitute the McdAB system for α-carboxysomes in Proteobacteria is of great interest for understanding the evolution, diversity, and mechanisms of carboxysome positioning across the bacterial world.

The *S. elongatus* McdAB System Might Have a Unique Origin

The phylogenic placement of *S. elongatus* and Type 1 McdAB systems relative to Type 2 systems suggest the latter is more ancestral. *Gloeobacter kilaueensis* JS1 is one of the most primitive species among living cyanobacteria due to the lack of thylakoid membranes (Rippka et al. 1974; Guglielmi et al. 1981; Nakamura et al. 2003). It possesses the widely distributed Type 2 system, suggesting that the Type 1 system of *S. elongatus* was acquired more recently. We obtained strong bootstrap support for *S. elongatus* sharing an immediate common ancestor with α-cyanobacteria, which lack the McdAB system, and four cyanobacteria where the McdAB system could not be identified (fig. 3). Many Type 1 McdA homologs identified here by BLASTp are plasmid encoded (the sole ParA-type ATPase on the plasmid) and the N-terminal domain of Type 1 McdB's are nearly identical to the small coding sequences following these plasmid-encoded McdA-like proteins. Therefore, the most parsimonious hypothesis is that the plasmid encoded ParA-type system used to partition plasmids was horizontally transferred into *S. elongatus*, genomically integrated, and evolved to position carboxysomes. A similar evolutionary path would explain the positioning of a diversity of cargos by ParA family ATPases (Lutkenhaus 2012; Vecchiarelli et al. 2012; Kiekebusch and Thanbichler 2014).

As the Type 2 system was present at the earliest known point of cyanobacterial evolution, it is not clear why the Type 1 system would have had a selective advantage over the Type 2 system, presumably present in the ancestors of *S. elongatus*. It is also interesting that we were unable to identify the McdAB system in several modern cousins of *S. elongatus* (fig. 3). An alternative hypothesis could be that this clade of β-cyanobacteria (which now includes modern *S. elongatus*) lost the Type 2 system, similar to the β-cyanobacterial clades found on the right side of our tree, and that *S. elongatus* then acquired the Type 1 system. As more cyanobacterial genomes are sequenced, how the Type 1 McdAB system evolved will become more apparent.

McdB Is a Phase Separating Protein

Our finding that McdB undergoes phase separation in vitro is intriguing on multiple fronts. First, two recent studies have demonstrated that α- and β-carboxysome formation involves LLPS (Oltrogge et al. 2019; Wang et al. 2019). With β-carboxysomes, the protein CcmM aggregates RuBisCO to form a procarboxysome (Cameron et al. 2013). CcmM exists in two forms: 1) full-length CcmM (58 kDa), which contains a carbonic anhydrase-like domain followed by three RuBisCO small subunit-like domains separated by flexible linkers, and 2) short-form CcmM (35 kDa), which lacks the carbonic anhydrase-like domain (Long et al. 2010). Binding of small subunit-like domain regions of CcmM between RbcL dimers scaffolds RuBisCO molecules and induces LLPS (Wang et al. 2019). Likewise, in α-carboxysomes of the chemoautotrophic proteobacterium *Halothiobacillus neapolitanus*, N-terminal repeated motifs within the intrinsically disordered protein CsoS2 mediate interaction with RuBisCO and induce LLPS (Cai et al. 2015; Oltrogge et al. 2019). The pyrenoid does not have a protein shell, but serves as the functional equivalent to carboxysomes in eukaryotic algae. One study of the pyrenoid showed that its RuBisCO matrix also exhibits liquid-like properties when interacting with the intrinsically disordered protein EPYC1 in *Chlamydomonas reinhardtii* (Freeman Rosenzweig et al. 2017). Collectively, these studies suggest that LLPS is a common feature underlying the formation of a RuBisCO matrix through interactions with intrinsically disordered proteins. Given that McdB is also an intrinsically disordered protein that undergoes LLPS (fig. 5*B–G* and supplementary videos 1 and 2, Supplementary Material online) and our previous data that increased levels of McdB drastically increase carboxysome size (MacCready et al. 2018), it is intriguing to speculate that the LLPS activities of McdB and the carboxysome core are related and potentially influence each other. It is possible that the McdAB system not only positions carboxysomes in the cell but also maintains homeostasis of a liquid-like carboxysome core.

Our results that *S. elongatus* McdB formed droplets across a broad pH range are informative. It has been previously suggested that a pH gradient exists between the cytosol and carboxysomes (Menon et al. 2010; Whitehead et al. 2014). Experimental evidence in *S. elongatus* showed that while the pH of the cytosol of *S. elongatus* is ~8.4 under light conditions, the carboxysome is predicted to be more acidic (pH ~6.0–7.0) (Mangan et al. 2016). An acidic carboxysome has been suggested to increase the maximum carboxylation rate of RuBisCO and reduce the amount of HCO_3^- uptake required to saturate RuBisCO (Mangan et al. 2016). Our in vitro data suggest that McdB would be soluble in the cytosol (pH $\geq$8) (fig. 5*B*) and undergo LLPS at or near acidic carboxysomes (pH $\leq$ 7.5) (fig. 5*B*). However, dark adapted cells of *S. elongatus* have a cytosolic pH ~7.3 (Mangan et al. 2016), so McdB behavior might differ in cells under light and dark conditions. Moreover, cyanobacteria possess a biological circadian clock that precisely operates on the 24-h rotational period of the earth. Circadian rhythms primarily enable cyanobacteria to anticipate, adapt, and respond to daily light

cycles by translating environmental cues into changes in gene expression (reviewed in Cohen and Golden 2015). In *S. elongatus*, oscillatory patterns of gene expression are driven by phosphorylation of the master output transcriptional regulator protein RpaA. Phosphorylation of RpaA has previously been shown to bind ~170 promoters of the *S. elongatus* chromosome (Markson et al. 2013); one of which is the promoter for the *mcdAB* operon. Therefore, it will be interesting to explore the role of McdB LLPS activity at carboxysomes, and how circadian rhythms and light-dark conditions influence McdAB expression, dynamics, and function.

Lastly, McdB is the first example of a ParA partner protein shown to have LLPS activity. It is attractive to speculate that under appropriate conditions, other ParA partner proteins exhibit similar LLPS behaviors that are critical to their in vivo function. The most obvious examples are ParB proteins that bind to a centromere-like DNA sequence called *parS* and are involved in plasmid and chromosome segregation in bacteria (reviewed in Wang et al. 2013; Baxter and Funnell 2014; Bouet et al. 2014). Through ChIP approaches, it is well known that thousands of ParB dimers associate with broad regions of DNA adjacent to the *parS* site, a phenomenon known as "spreading" (Rodionov et al. 1999; Murray et al. 2006; Breier and Grossman 2007; Sanchez et al. 2015; Debaugny et al. 2018). However, the actual structure of the ParB-DNA mega-complex remains unclear. In vivo, fluorescent fusions of ParB form a massive punctate focus at the location of a *parS* site on the chromosome or plasmid. These ParB foci segregate after DNA replication in a ParA-dependent manner and have also been shown to undergo rapid fission/fusion events, whereby ParB foci segregate and then snap back together (Sengupta et al. 2010). FRAP measurements of ParB exchange at these foci also suggest these complexes are highly dynamic (Debaugny et al. 2018). Together, the data are consistent with ParB forming a biomolecular condensate with its *parS* site and its adjacent DNA. LLPS activity of ParB proteins on *parS*-containing DNA substrates has yet to be directly observed in vitro. However, in silico models have proposed that the currently known affinities of ParB–ParB and ParB–DNA associations could induce the formation of a condensate (Broedersz et al. 2014; Sanchez et al. 2015; Debaugny et al. 2018). The vast majority of proteins capable of LLPS do so by interacting with DNA/RNA (reviewed in Alberti et al. 2019). McdB, however, does not bind DNA and is capable of LLPS activity on its own. A future direction of research is aimed at understanding how the LLPS activity of McdB, and other ParA partner proteins, is involved in the spatial organization of their cognate cargos.

McdB Charge Distribution Might Contribute to LLPS

McdB proteins across evolutionary time possess many of the features that enable LLPS including intrinsic disorder, low hydrophobicity, biased amino acid compositions, and extreme multivalency (reviewed in Alberti et al. 2019) (fig. 4). When analyzing the multivalent properties of McdB proteins, we found that the vast majority was polyampholytes with biphasic charge distributions between their N- and C-terminal extensions flanking the coiled-coil domain (fig. 5*K* and supplementary fig. S4, Supplementary Material online). The reason for such a shared feature is not obvious. Although genetic drift possibly accounts for the intrinsic disorder and lack of primary sequence conservation among McdB proteins, it is unlikely that genetic drift alone could account for the patterned charge distributions. Indeed, given that the vast majority of McdB proteins exhibit an inverted charged patterning along their N- and C-terminal extensions (~68%, fig. 5*K* and supplementary fig. S4, Supplementary Material online), a more parsimonious explanation could be that the charged extensions are important for McdB associations with itself or with carboxysome shell proteins. Although not a true McdB protein, the structure of the coiled-coil domain of a plasmid-encoded McdB-like protein from the cyanobacterium *Cyanothece* sp. PCC 7424 (PCC7424_5530) demonstrated an antiparallel association to form a dimer (Schumacher et al. 2019). Parallel or antiparallel assembly of the coiled-coil domains of McdB would either align like-or oppositely charged extensions. The role of multivalency, charge patterning, and dimerization on the LLPS activity of McdB are interesting avenues of future investigation.

mcdA and *mcdB* Genes Often Cluster with the Minor Carboxysome Shell Genes CcmK3 and CcmK4

Of the *mcdA* and *mcdB* genes, we found near carboxysome components, the majority clustered near the minor shell genes *ccmK3* and *ccmK4* (figs. 1*E* and 2*A*); two shell genes often found together, but at a different locus than the carboxysome operon (Sommer et al. 2017). The finding is consistent with our prior results showing that McdB interacts with CcmK3 and CcmK4 in a bacterial two-hybrid system (MacCready et al. 2018). Moreover, carboxysomes have been shown to cluster following the deletion of CcmK3 and CcmK4 in a manner that is reminiscent of carboxysome aggregation in our Δ*mcdB* strain (Rae et al. 2013; MacCready et al. 2018). How McdB and CcmK3/CcmK4 interact is still an open question. A recent study showed that CcmK3 and CcmK4 can form homohexamers, as well as heterohexamers that further assemble into dodecamers under certain conditions (Sommer et al. 2019). Metabolite channeling into and out of carboxysomes is believed to occur via the central pores of these hexameric shell proteins (Dou et al. 2008; Kinney et al. 2011). As the pore formed by CcmK4 homohexamers differs from those formed by CcmK3/CcmK4 heterohexamers, it is proposed that homo- and hetero-hexamic species alter carboxysome permeability, thereby modulating metabolite channeling across the protein shell (Sommer et al. 2019). Interestingly, CcmK3/CcmK4 dodecamer formation was found to be influenced by pH (relative dodecamer abundance of 36% at pH 7.0 and 3% at pH 8) (Sommer et al. 2019). Likewise, we find here that McdB LLPS behavior is also greatly influenced by transitions in this pH range; soluble at a pH $\geq$8 and formed liquid droplets at a pH $\leq$ 7.5 (fig. 5*B*). Placing these findings in the context of current models that propose the carboxysome as more acidic (pH ~ 7.0) than the cytosol (pH ~ 8.4) in *S. elongatus* (Menon et al. 2010; Whitehead et al. 2014; Mangan et al. 2016), several important questions are

revealed. How does McdB interact with CcmK3/CcmK4 assemblies? Is this interaction influenced by McdB LLPS activity and by the differential pH of the carboxysome versus the cytoplasm? Finally, does McdB influence metabolite channeling into and out of carboxysomes and is this activity dependent on LLPS behavior of McdB? It is attractive to speculate that local pH changes due to the metabolic activities of the carboxysomes also serves to regulate its composition, structure, and function. Together, our results have broad implication for understanding the diversity of the McdAB system, the mechanisms of carboxysome positioning among carbon-fixing bacteria, and the role of LLPS in the biogenesis and spatial organization of carboxysomes.

Materials and Methods

McdAB Homolog Search and Neighborhood Analysis

Initial searches for *S. elongatus* McdA (Synpcc7942_1833) and McdB (Synpcc7942_1834) homologs were performed via BLASTp. As BLASTp returned few hits for McdA and not a single reliable hit for McdB, we reasoned that neighborhood analysis was a better method for identifying possible homologs of these proteins. Neighborhood analysis was performed by obtaining genome assembly GenBank files from NCBI for all cyanobacteria used in this study. As genome annotations are inconsistent among cyanobacterial species and carboxysome components can be found across multiple loci, homologs of *S. elongatus* carboxysome components CcmK2 (Synpcc7942_1421), CcmK3 (Synpcc7942_0284), CcmK4 (Synpcc7942_0285), CcmL (Synpcc7942_1422), CcmM (Synpcc7942_1423), CcmN (Synpcc7942_1424), CcmO (Synpcc7942_1425), CcmP (Synpcc7942_0520), RbcS (Synpcc7942_1427), RbcL (Synpcc7942_1426), RbcX (Synpcc7942_1535), CcaA (Synpcc7942_1447), and RbcR (Synpcc7942_1980) were used as BLASTp queries to identify carboxysome components in all other cyanobacteria. We note, additional shell proteins exist among other cyanobacteria, including CcmK1, CcmK5, and CcmK6 (Sommer et al. 2017). However, these proteins share high similarity to CcmK2, so they were also captured during our BLASTp search using CcmK2.

Neighborhood analysis for each carboxysome component across all cyanobacterial genomes was performed manually. We defaulted to this approach as many cyanobacterial genomes were still drafts and we wanted to determine whether the contigs or scaffolds were too small to make an appropriate neighborhood determination. Moreover, we also wanted the ability to quantify the number of base-pairs and coding sequences between *mcdA* or *mcdB* and the gene of the neighboring carboxysome component. Neighborhood analysis was carried out via Biomatters Geneious v 11.1.5 by searching 10 kb up- and downstream of each carboxysome component gene across all cyanobacterial species used in this study to identify coding sequences that matched our criteria for McdA or McdB. Accession numbers of identified McdA, McdB, and carboxysome component(s) were recorded using Excel.

McdAB Sequence Analysis

Multiple sequence alignments for McdA proteins were performed using MAFFT 1.3.7 under the G-INS-I algorithm, whereas the E-INS-I algorithm was used for McdB proteins due to long gaps caused by intrinsic disorder. Coiled-coil predictions for McdB proteins were carried out using DeepCoil (Ludwiczak et al. 2019). Predictions of disorder within McdB proteins were performed using PONDR with the VL-XT algorithm (Romero et al. 1997, 2001; Li et al. 1999). Analysis of McdB hydrophobicity was conducted with ProtScale using the Kyte and Doolittle scale (Kyte and Doolittle 1982; Gasteiger et al. 2005).

Phylogenetic Inference

Ortholog sequences for *S. elongatus* DnaG, RplA, RplB, RplC, RplD, and RplE were obtained via BLASTp for each cyanobacterium. Alignments for protein sequence were performed using MAFFT 1.3.7 under the G-INS-I algorithm and BLOSUM62 scoring matrix. The six resulting alignments were then concatenated into one alignment using Geneious 11.1.5. Regions of low conservation within the resulting alignment were removed using gBlocks 0.91 b (Castresana 2000; Talavera and Castresana 2007). A phylogenetic tree was then estimated with maximum likelihood analyses using RAxML 8.2.11 under the LG+Gamma scoring model of amino acid substitution. Bootstrap values were calculated from 500 replicates. The resulting phylogenetic tree was then edited and visualized using iTOL3 (Letunic and Bork 2016).

Construct Development

The His–SUMO–McdB expression plasmid was generated via Gibson Assembly (Gibson et al. 2009) using synthetized dsDNA for His–SUMO–McdB that was inserted into a pET11b vector. Construct was verified by sequencing. Cloning was performed in *Escherichia coli* DH5α chemically competent cells (Invitrogen).

His–SUMO–McdB Overexpression

The His–SUMO–McdB fusion was expressed from a pET11b vector containing an ampicillin resistance cassette. All *E. coli* cultures containing the vector were grown in LB + carbenicillin (100 μg/ml) at 37 °C with 220 rpm shaking unless otherwise stated. The vector was transformed into competent *E. coli* BL21-AI, and a single transformant was picked to inoculate a 25 ml overnight culture. For expression, 2 l of culture was inoculated through a 1:100 dilution of the overnight culture and grown at 37 °C until an OD_{600} of 0.5–0.7 was reached. Expression was induced with final concentrations of IPTG at 1 mM and L-arabinose at 0.2%. Immediately after induction, cultures were cooled in an ice bath for 5 min and shifted to 18 °C to grow for 16 h at 220 rpm. Cells were pelleted at 4000×g for 20 min at 4 °C, flash frozen in liq. N_2, and stored at −80 °C.

His–SUMO–McdB Purification

The following buffers were used to purify the His–SUMO–McdB fusion: lysis buffer (50 mM HEPES pH 7.6; 1 M KCl; 5 mM $MgCl_2$; 10% glycerol; 2 mM BME; 20 mM imidazole;

0.05 mg/ml lysozyme; 0.05 µl/ml benzonase; protease inhibitor), Ni buffer A (50 mM HEPES pH 7.6; 1 M KCl; 5 mM $MgCl_2$; 10% glycerol; 2 mM BME; 20 mM imidazole), and Ni buffer B (50 mM HEPES pH 7.6; 1 M KCl; 5 mM $MgCl_2$; 10% glycerol; 2 mM BME; 500 mM imidazole). The 2-l cell pellet was resuspended in 150 ml lysis buffer and lysed using a microfluidizer at 18,000 psi equilibrated in Ni buffer A. The lysate was centrifuged at 30,000×g at 4 °C for 30 min. The clarified lysate was decanted, syringe filtered (0.2 μm), and loaded onto a 5 ml HP HisTRAP column equilibrated in Ni buffer A. The column was washed with 25 ml Ni buffer A and then 25 ml 5% Ni buffer B. Elution was done using a 5–100% gradient of Ni buffer B while collecting 2 ml fractions for a total of 60 ml. Peak fractions were flash frozen in N_2 (liq.) and stored at −80 °C.

Microscopy of In Vitro McdB Phase Separation

Prior to imaging, His–SUMO–McdB samples from *S. elongatus* 7942 were buffer exchanged into a Ulp1 reaction buffer (150 mM KCl; 25 mM HEPES; 2 mM BME) and adjusted to the indicated pH. Similarly, His–SUMO–McdB samples from each of the Type 2 systems chosen were buffer exchanged into a defined Ulp1 reaction buffer, being *Gloeobacter kilaueensis* JSI (150 mM NaCl; 25 mM HEPES, pH 7.0; 2 mM BME), *Fremyella diplosiphon* NIES-3275 (25 mM HEPES, pH 7.0; 2 mM BME), and *Fischerella* sp. PCC 9431 (150 mM NaCl; 25 mM HEPES, pH 7.0; 2 mM BME). Buffer exchange was performed using 7 K MWCO, 5 ml Zeba Spin Desalting Columns (Thermo-Fischer). All imaging was performed using 16 well CultureWells (Grace BioLabs). Wells were passivated by overnight incubation in 5% (w/v) Pluronic acid (Thermo-Fischer), and washed thoroughly with the corresponding Ulp1 buffer prior to use. For cleavage experiments, 1 µl of purified Ulp1 was added to 50 µl of His–SUMO–McdB at the indicated concentration and incubated at 23 °C for 2 h to ensure complete cleavage. Imaging of McdB droplet formation was performed using a Nikon Ti2-E motorized inverted microscope (60× DIC objective and DIC analyzer cube) with a Transmitted LED Lamp house and a Photometrics Prime 95B Back-illuminated sCMOS Camera. Image analysis was performed using Fiji v 1.0.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank Lindsay Matthews from the Simmons Lab for providing the Ulp1 enzyme. This work was supported by the National Science Foundation to A.G.V. (Award No. 1817478 and CAREER Award No. 1941966), research initiation funds provided by the MCDB Department to A.G.V., University of Michigan, and by research funds from the Michigan Life Sciences Fellows Program to J.S.M.

Author Contributions

J.S.M. and A.G.V. conceived the project. J.S.M., J.L.B., and A.G.V. designed experiments. J.S.M. and J.L.B. performed all experiments. J.S.M. and A.G.V. wrote the article. All authors discussed results and edited the article.

References

Adachi S, Hori K, Hiraga S. 2006. Subcellular positioning of F plasmid mediated by dynamic localization of SopA and SopB. *J Mol Biol*. 356(4):850–863.

Ah-Seng Y, Lopez F, Pasta F, Lane D, Bouet JY. 2009. Dual role of DNA in regulating ATP hydrolysis by the SopA partition protein. *J Biol Chem*. 284(44):30067–30075.

Alberti S, Gladfelter A, Mittag T. 2019. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* 176(3):419–434.

Alvarado A, Kjær A, Yang W, Mann P, Briegel A, Waldor MK, Ringgaard S. 2017. Coupling chemosensory array formation and localization. *eLife* 6:e31058.

Apostolovic B, Danial M, Klok HA. 2010. Coiled coils: attractive protein folding motifs for the fabrication of self-assembled, responsive and bioactive materials. *Chem Soc Rev*. 39(9):3541–3575.

Atmakuri K, Cascales E, Burton OT, Banta LM, Christie PJ. 2007. Agrobacterium ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions. *EMBO J*. 26(10): 2540–2551.

Axen SD, Erbilgin O, Kerfeld CA. 2014. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol*. 10(10):e1003898.

Badrinarayanan A, Le TB, Laub MT. 2015. Bacterial chromosome organization and segregation. *Annu Rev Cell Dev Biol*. 31(1): 171–199.

Barillà D, Carmelo E, Hayes F. 2007. The tail of the ParG DNA segregation protein remodels ParF polymers and enhances ATP hydrolysis via an arginine finger-like motif. *Proc Natl Acad Sci U S A*. 104(6):1811–1816.

Baxter JC, Funnell BE. 2014. Plasmid partition mechanisms. *Microbiol Spectr*. 2(6): PLAS-0023-2014.

Beck C, Knoop H, Steuer R. 2018. Modules of co-occurrence in the cyanobacterial pan-genome reveal functional associations between groups of ortholog genes. *PLoS Genet*. 14(3):e1007239.

Bouet JY, Funnell BE. 1999. P1 ParA interacts with the P1 partition complex at parS and an ATP-ADP switch controls ParA activities. *EMBO J*. 18(5):1415–1424.

Bouet J-Y, Stouf M, Lebailly E, Cornet F. 2014. Mechanisms for chromosome segregation. *Curr Opin Microbiol*. 22:60–65.

Bouet JY, Surtees JA, Funnell BE. 2000. Stoichiometry of P1 plasmid partition complexes. *J Biol Chem*. 275(11):8213–8219.

Breier AM, Grossman AD. 2007. Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the *Bacillus subtilis* chromosome. *Mol Microbiol*. 64(3):703–718.

Broedersz CP, Wang X, Meir Y, Loparo JJ, Rudner DZ, Wingreen NS. 2014. Condensation and localization of the partitioning protein ParB on the bacterial chromosome. *Proc Natl Acad Sci U S A*. 111(24):8809–8814.

Cai F, Dou Z, Bernstein SL, Leverenz R, Williams EB, Heinhorst S, Shively J, Cannon GC, Kerfeld CA. 2015. Advances in understanding carboxysome assembly in *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component. *Life* 5(2):1141–1171.

Cameron JC, Wilson SC, Bernstein SL, Kerfeld CA. 2013. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* 155(5):1131–1140.

Castaing JP, Bouet JY, Lane D. 2008. F plasmid partition depends on interaction of SopA with non-specific DNA. *Mol Microbiol*. 70(4):1000–1011.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17(4):540–552.
Cohen SE, Golden SS. 2015. Circadian rhythms in cyanobacteria. *Microbiol Mol Biol Rev*. 79(4):373–385.
Davis MA, Austin SJ. 1988. Recognition of the P1 plasmid centromere analog involves binding of the ParB protein and is modified by a specific host factor. *EMBO J*. 7(6):1881–1888.
Davis MA, Martin KA, Austin SJ. 1992. Biochemical activities of the parA partition protein of the P1 plasmid. *Mol Microbiol*. 6(9):1141–1147.
Debaugny RE, Sanchez A, Rech J, Labourdette D, Dorignac J, Geniet F, Palmeri J, Parmeggiani A, Boudsocq F, Anton Leberre V, et al. 2018. A conserved mechanism drives partition complex assembly on bacterial chromosomes and plasmids. *Mol Syst Biol*. 14(11):e8516.
Deeb SS. 1973. Involvement of a tryptophan residue in the assembly of bacteriophages 80 and lambda. *J Virol*. 11(3):353–358.
Dou Z, Heinhorst S, Williams EB, Murin CD, Shively JM, Cannon GC. 2008. CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. *J Biol Chem*. 283(16):10377–10384.
Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CC, Eckmann CR, Myong S, Brangwynne CP. 2015. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc Natl Acad Sci U S A*. 112(23):7189–7194.
Freeman Rosenzweig ES, Xu B, Kuhn Cuellar L, Martinez-Sanchez A, Schaffer M, Strauss M, Cartwright HN, Ronceray P, Plitzko JM, Förster F, et al. 2017. The eukaryotic CO2-concentrating organelle Is liquid-like and exhibits dynamic reorganization. *Cell* 171(1):148–162.e19.
Funnell BE. 1988. Mini-P1 plasmid partitioning: excess ParB protein destabilizes plasmids containing the centromere parS. *J Bacteriol*. 170(2):954–960.
Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The proteomics protocols handbook. Totowa, NJ: Humana Press Inc., p. 571–607.
Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 6(5):343–345.
Gonzalez-Esquer CR, Smarda J, Rippka R, Axen SD, Guglielmi G, Gugger M, Kerfeld CA. 2016. Cyanobacterial ultrastructure in light of genomic sequence data. *Photosyn Res*. 129(2):147–157.
Griese M, Lange C, Soppa J. 2011. Ploidy in cyanobacteria. *FEMS Microbiol Lett*. 323(2):124–131.
Guglielmi G, Cohen-Bazire G, Bryant DA. 1981. The structure of *Gloeobacter violaceus* and its phycobilisomes. *Arch Microbiol*. 129(3):181–189.
Harmon TS, Holehouse AS, Rosen MK, Pappu RV. 2017. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* 6:pii:e30294.
Hatano T, Yamaichi Y, Niki H. 2007. Oscillating focus of SopA associated with filamentous structure guides partitioning of F plasmid. *Mol Microbiol*. 64(5):1198–1213.
Hester CM, Lutkenhaus J. 2007. Soj (ParA) DNA binding is mediated by conserved arginines and is essential for plasmid segregation. *Proc Natl Acad Sci U S A*. 104(51):20326–20331.
Hirose Y, Chihong S, Watanabe M, Yonekawa C, Murata K, Ikeuchi M, Eki T. 2019. Diverse chromatic acclimation processes regulating phycoerythrocyanin and rod-shaped phycobilisome in cyanobacteria. *Mol Plant*. 12(8):1167–1169.
Hu L, Vecchiarelli AG, Mizuuchi K, Neuman KC, Liu J. 2017. Brownian ratchet mechanism for faithful segregation of low-copy-number plasmids. *Biophys J*. 112(7):1489–1502.
Hwang LC, Vecchiarelli AG, Han YW, Mizuuchi M, Harada Y, Funnell BE, Mizuuchi K. 2013. ParA-mediated plasmid partition driven by protein pattern self-organization. *EMBO J*. 32(9):1238–1249.
Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC, Jensen GJ. 2010. Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *J Mol Biol*. 396(1):105–117.
Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A, Parker R. 2016. ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell* 164(3):487–498.
Kaneko Y, Danev R, Nagayama K, Nakamoto H. 2006. Intact carboxysomes in a cyanobacterial cell visualized by Hilbert differential contrast transmission electron microscopy. *J Bacteriol*. 188(2):805–808.
Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, et al. 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149(4):753–767.
Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M. 2018. Bacterial microcompartments. *Nat Rev Microbiol*. 16(5):277–290.
Kiekebusch D, Thanbichler M. 2014. Plasmid segregation by a moving ATPase gradient. *Proc Natl Acad Sci U S A*. 111(13):4741–4742.
Kinney JN, Axen SD, Kerfeld CA. 2011. Comparative analysis of carboxysome shell proteins. *Photosynth Res*. 109(1-3):21–32.
Komla-Soukha I, Sureau C. 2006. A tryptophan-rich motif in the carboxyl terminus of the small envelope protein of hepatitis B virus is central to the assembly of hepatitis delta virus particles. *J Virol*. 80(10):4648–4655.
Kusumoto A, Shinohara A, Terashima H, Kojima S, Yakushi T, Homma M. 2008. Collaboration of FlhF and FlhG to regulate polar-flagella number and localization in *Vibrio alginolyticus*. *Microbiology* 154(5):1390–1399.
Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 157(1):105–132.
Lechno-Yossef S, Rohnke BA, Belza ACO, Melnicki MR, Montgomery BL, Kerfeld CA. 2019. Cyanobacterial carboxysomes contain an unique RuBisCO-activase-like protein. *New Phytol*. 225(2):793–806.
Leonard TA, Butler PJ, Lowe J. 2005. Bacterial chromosome segregation: structure and DNA binding of the Soj dimer–a conserved biological switch. *EMBO J*. 24(2):270–282.
Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 44(W1):W242–W245.
Li P, Banjade S, Cheng HC, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth JV, King DS, Banani SF, et al. 2012. Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483(7389):336–340.
Li X, Romero P, Rani M, Dunker AK, Obradovic Z. 1999. Predicting protein disorder for N-, C-, and internal regions. *Genome Informatics*. 10:30–40.
Lin Y, Protter DS, Rosen MK, Parker R. 2015. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol Cell*. 60(2):208–219.
Long BM, Tucker L, Badger MR, Price GD. 2010. Functional cyanobacterial beta-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiol*. 153(1):285–293.
Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. 2019. DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics* 35(16):2790–2795.
Lutkenhaus J. 2012. The ParA/MinD family puts things in their place. *Trends Microbiol*. 20(9):411–418.
MacCready JS, Hakim P, Young EJ, Hu L, Liu J, Osteryoung KW, Vecchiarelli AG, Ducat DC. 2018. Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* 7:pii:e39723.
MacCready JS, Schossau J, Osteryoung KW, Ducat DC. 2017. Robust Min-system oscillation in the presence of internal photosynthetic membranes in cyanobacteria. *Mol Microbiol*. 103(3):483–503.
Mangan NM, Flamholz A, Hood RD, Milo R, Savage DF. 2016. pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc Natl Acad Sci U S A*. 113(36):E5354–E5362.
Marin B, Nowack EC, Glöckner G, Melkonian M. 2007. The ancestor of the *Paulinella* chromatophore obtained a carboxysomal operon by horizontal gene transfer from a Nitrococcus-like gamma-proteobacterium. *BMC Evol Biol*. 7(1):85.

Marintcheva B, Hamdan SM, Lee SJ, Richardson CC. 2006. Essential residues in the C terminus of the bacteriophage T7 gene 2.5 single-stranded DNA-binding protein. *J Biol Chem*. 281(35):25831–25840.

Markson JS, Piechura JR, Puszynska AM, O'Shea EK. 2013. Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* 155(6):1396–1408.

Menon BB, Heinhorst S, Shively JM, Cannon GC. 2010. The carboxysome shell is permeable to protons. *J Bacteriol*. 192(22):5881–5886.

Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, Mittag T, Taylor JP. 2015. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163(1):123–133.

Montgomery BL. 2015. Light-dependent governance of cell shape dimensions in cyanobacteria. *Front Microbiol*. 6:514.

Murray H, Ferreira H, Errington J. 2006. The bacterial chromosome segregation protein Spo0J spreads along DNA from parS nucleation sites. *Mol Microbiol*. 61(5):1352–1361.

Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, et al. 2003. Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Res*. 10(4):137–145.

Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, et al. 2015. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell*. 57(5):936–947.

Oldfield CJ, Dunker AK. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem*. 83(1):553–584.

Oliva MA, Martin-Galiano AJ, Sakaguchi Y, Andreu JM. 2012. Tubulin homolog TubZ in a phage-encoded partition system. *Proc Natl Acad Sci U S A*. 109(20):7711–7716.

Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S, Savage DF. 2019. α-Carboxysome formation is mediated by the multivalent and disordered protein CsoS2. BioRxiv. doi: https://doi.org/10.1101/708164.

Perez-Cheeks BA, Planet PJ, Sarkar IN, Clock SA, Xu Q, Figurski DH. 2012. The product of tadZ, a new member of the parA/minD superfamily, localizes to a pole in *Aggregatibacter actinomycetemcomitans*. *Mol Microbiol*. 83(4):694–711.

Radnedge L, Youngren B, Davis M, Austin S. 1998. Probing the structure of complex macromolecular interactions by homolog specificity scanning: the P1 and P7 plasmid partition systems. *EMBO J*. 17(20):6076–6085.

Rae BD, Long BM, Badger MR, Price GD. 2013. Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol R*. 77(3):357–379.

Raskin DM, de Boer PA. 1999. Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*. *Proc Natl Acad Sci U S A*. 96(9):4971–4976.

Ravin NV, Rech J, Lane D. 2003. Mapping of functional domains in F plasmid partition proteins reveals a bipartite SopB-recognition domain in SopA. *J Mol Biol*. 329(5):875–889.

Ringgaard S, Schirner K, Davis BM, Waldor MK. 2011. A family of ParA-like ATPases promotes cell pole maturation by facilitating polar localization of chemotaxis proteins. *Gene Dev*. 25(14): 1544–1555.

Rippka R, Waterbury J, Cohen-Bazire G. 1974. A cyanobacterium which lacks thylakoids. *Arch Microbiol*. 100(1):419–436.

Rodionov O, Lobocka M, Yarmolinsky M. 1999. Silencing of genes flanking the P1 plasmid centromere. *Science* 283(5401):546–549.

Romero P, Obradovic Z, Dunker AK. 1997. Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Informatics*. 8: 110–124.

Romero P, Obradovic Z, Li X, Garner E, Brown C, Dunker AK. 2001. Sequence complexity of disordered protein. *Proteins* 42(1):38–48.

Sanchez A, Cattoni DI, Walter JC, Rech J, Parmeggiani A, Nollmann M, Bouet JY. 2015. Stochastic self-assembly of ParB proteins builds the bacterial DNA segregation apparatus. *Cell Syst*. 1(2):163–173.

Savage DF, Afonso B, Chen AH, Silver PA. 2010. Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* 327(5970):1258–1261.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I, Partensky F. 2009. Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev*. 73(2):249–299.

Schumacher MA, Henderson M, Zhang H. 2019. Structures of maintenance of carboxysome distribution Walker-box McdA and McdB adaptor homologs. *Nucleic Acids Res*. 47(11):5950–5962.

Schuster BS, Reed EH, Parthasarathy R, Jahnke CN, Caldwell RM, Bermudez JG, Ramage H, Good MC, Hammer DA. 2018. Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat Commun*. 9(1):2985.

Sengupta M, Nielsen HJ, Youngren B, Austin S. 2010. P1 plasmid segregation: accurate redistribution by dynamic plasmid pairing and separation. *J Bacteriol*. 192(5):1175–1183.

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, et al. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A*. 110(3):1053–1058.

Shin Y, Berry J, Pannucci N, Haataja MP, Toettcher JE, Brangwynne CP. 2017. Spatiotemporal control of intracellular phase transitions using light-activated optoDroplets. *Cell* 168(1–2):159–171.e14.

Sommer M, Cai F, Melnicki M, Kerfeld CA. 2017. β-Carboxysome bioinformatics: identification and evolution of new bacterial microcompartment protein gene classes and core locus constraints. *J Exp Bot*. 68(14):3841–3855.

Sommer M, Sutter M, Gupta S, Kirst H, Turmo A, Lechno-Yossef S, Burton RL, Saechao C, Sloan NB, Cheng X, et al. 2019. Heterohexamers formed by CcmK3 and CcmK4 increase the complexity of beta carboxysome shells. *Plant Physiol*. 179(1):156–167.

Sun Y, Casella S, Fang Y, Huang F, Faulkner M, Barrett S, Liu LN. 2016. Light modulates the biosynthesis and organization of cyanobacterial carbon fixation machinery through photosynthetic electron flow. *Plant Physiol*. 171(1):530–541.

Sun Y, Wollman AJM, Huang F, Leake MC, Liu LN. 2019. Single-organelle quantification reveals stoichiometric and structural variability of carboxysomes dependent on the environment. *Plant Cell*. 31(7):1648–1664.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 56(4):564–577.

Thompson SR, Wadhams GH, Armitage JP. 2006. The positioning of cytoplasmic protein clusters in bacteria. *Proc Natl Acad Sci U S A*. 103(21):8209–8214.

Toro-Nahuelpan M, Müller FD, Klumpp S, Plitzko JM, Bramkamp M, Schüler D. 2016. Segregation of prokaryotic magnetosomes organelles is driven by treadmilling of a dynamic actin-like MamK filament. *BMC Biol*. 14(1):88.

Tsuboi M, Overman SA, Nakamura K, Rodriguez-Casado A, Thomas GJ Jr. 2003. Orientation and interactions of an essential tryptophan (Trp-38) in the capsid subunit of Pf3 filamentous virus. *Biophys J*. 84(3):1969–1976.

Varadi M, Zsolyomi F, Guharoy M, Tompa P. 2015. Functional advantages of conserved intrinsic disorder in RNA-binding proteins. *PLoS One* 10(10):e0139731.

Vecchiarelli AG, Han YW, Tan X, Mizuuchi M, Ghirlando R, Biertumpfel C, Funnell BE, Mizuuchi K. 2010. ATP control of dynamic P1 ParA–DNA interactions: a key role for the nucleoid in plasmid partition. *Mol Microbiol*. 78(1):78–91.

Vecchiarelli AG, Mizuuchi K, Funnell BE. 2012. Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Mol Microbiol*. 86(3):513–523.

Vecchiarelli AG, Neuman KC, Mizuuchi K. 2014. A propagating ATPase gradient drives transport of surface-confined cellular cargo. *Proc Natl Acad Sci U S A*. 111(13):4880–4885.

Viollier PH, Sternheim N, Shapiro L. 2002. A dynamically localized histidine kinase controls the asymmetric distribution of polar pili proteins. *EMBO J*. 21(17):4420–4428.

Wang D, Seeve C, Pierson LS III, Pierson EA. 2013. Transcriptome profiling reveals links between ParS/ParR, MexEF-OprN, and quorum sensing in the regulation of adaptation and virulence in *Pseudomonas aeruginosa*. *BMC Genomics* 14:618.

Wang H, Yan X, Aigner H, Bracher A, Nguyen ND, Hee WY, Long BM, Price GD, Hartl FU, Hayer-Hartl M. 2019. RuBisCO condensate formation by CcmM in β-carboxysome biogenesis. *Nature* 566(7742):131–135.

Whitehead L, Long BM, Price GD, Badger MR. 2014. Comparing the in vivo function of α-carboxysomes and β-carboxysomes in two model cyanobacteria. *Plant Physiol*. 165(1):398–411.

Youderian P, Burke N, White DJ, Hartzell PL. 2003. Identification of genes required for adventurous gliding motility in *Myxococcus xanthus* with the transposable element mariner. *Mol Microbiol*. 49(2):555–570.